LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Bounded Generativity: Contextualising interdependencies between architecture, ecosystem, and environment in digital product innovation

Studying the design and development of an agent-based simulation model for forced displacement

Anulekha Nandi

A thesis submitted to the Information Systems and Innovation Faculty Group, Department of Management of the London School of Economics and Political Science for the degree of Doctor of Philosophy.

London, July 2024

DECLARATION

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 80,131 words.

ABSTRACT

Existing theorisation on digital product innovation remains predicated on a particular architectural form (modularity) and mode (unbounded generativity) of organising at scale participation of heterogeneous actors in an ecosystem. Despite the widely accepted role of product architectures in organising digital product innovation there has been limited academic engagement beyond the dynamics of modular design and its proximate context of the ecosystem. While contextualist research within information systems acknowledges the existence of wider systemic conditions underlying IS innovation, this has not received adequate attention within digital product innovation. This thesis builds on existing literature to understand the nature of interdependencies between the architecture, its proximate context of the ecosystem, and the distant context of the wider environment with the aim of developing a contextualised theory of digital product innovation for an alternative architectural form.

To augment and extend existing theory, this research studies the design and development of an agent-based simulation model for forced displacement. It uses Kleine's Choice Framework, adapted for this study, to understand how different conditions of possibility within the proximate and distant contexts shape operational and substantive choices within a digital product's ongoing development. It follows a process research approach to unpack the sequence of events, its constituent elements, and causal trajectories over time. It is based on an in-depth case study constructed through year-long field work with the development team along with the study of associated documents and reports. The research contributes to the theory on digital product innovation by unpacking how this trilateral interdependency creates opportunity structures at different stages of the development process which shape and bound the generative potential of digital products. This thesis demonstrates how this occurs through complementary resource-relationship configurations which negotiate the systemic conditions of multiple environmental drivers and technical conditions of a hybrid digital architecture.

ACKNOWLEDGEMENT

My heartfelt thanks to:
My supervisor, Dr. Shirin Madon for gentle and continuous guidance in shaping up this thesis from a mess of ideas to a coherent whole | Dr. Edgar Whitley, second supervisor and Programme Director; Dr. Carsten Sørensen, former Programme Director; Dr. Antonio Cordella and Dr. Jannis Kallinikos, upgrade examiners, for their invaluable advice at different stages of the thesis | The Research Operations team of Camilla Kennedy-Harper for navigating the PhD process, Emma Ward for stellar support with the organisation of the KEI event, and Nino Nizharadze for input and advice on the KEI funding application | The LSE Studentship without which I could not have undertaken the PhD | The LSE Knowledge Exchange and Impact Fund for the grant to organise the multistakeholder event on 'Emerging technologies in humanitarian management' | All event participants for taking out time from their extremely busy schedules to contribute their views and share their experiences | The BUL team for access, time, and patience for which I remain immensely grateful | My doctors for patching me up and supporting me through exigencies | Jane for making me feel at home, cats, Christmas lunch / dinner, our many chats, and listening to my rants | Toby for generous conversations, appreciation of a hearty meal, laughter, and being a wonderful and supportive friend | Nidhi for her kindness | Asheef for being a good friend even when I fail to be so many times | Ramyajit and Shunit for sheer madcap irreverence yet fortitude, resilience, and determination | Antara and Saswata Bose, their endearing little ones Aabir and Suchitra, and Noel for keeping up my spirits during this last leg | Shaku for holding my hand through difficult times, dispelling the gloom, food, and all round bracing and steadfast support; her family for many an evening of wholesome conversation | Grandparents for childhood treats, stories, and the love of Bengali literature | My brother, Abhilash, for showing how to bear many responsibilities seamlessly | My sister-in-law, Shreya, for being a refreshing change | The sheer joy and delight that are my niece and nephew, Gayatri and Jayveer | My parents, Mrinal and Jula, for giving us more than they ever had, not least the life opportunities with which one can fathom an advanced degree | Adhiraj for friendship and love over the years, thank you for keeping me grounded.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ABM - Agent Based Simulation Model

ACLED - Armed Conflict Location and Events Data Project

API - Application Programming Interface

BUL - Brunel University London

CoeGSS - Centre of Excellence for Global Systems Science

ComPat - Computing Patterns for High Performance Multiscale Computing

ECMWF - European Centre for Medium-Range Weather Forecasts

EMT - EUMigraTool

EU - European Union

FLP - Facility Location Problem

GDPR - General Data Protection Regulation

GloFAS - Global Flood Forecasting System

HiDALGO - HPC and Big Data Technologies for Global Challenges

HLRS - High Performance Computing Center Stuttgart

HPC - High Performance Computing

HPDA - High Performance Data Analytics

IDP - Internally Displaced Person

ICT - Information and Communication Technologies

ICT4D - Information and Communication Technologies for Development

IOM - International Organization for Migration

IPC - Integrated Food Security Phase Classification

IS - Information Systems

MOO - Multi Objective Optimisation

MPI4Py - Message Passing Interface for Python

MUSCLE - Multiscale Coupling Library and Environment

NGO - Non-Governmental Organisation

PSNC - Poznan Supercomputing and Networking Center

RFLP - Robust Facility Location Problem

RTDI - Recent Distance Travelled Index

SDA - Simulation Development Approach

SDSD - Sensitivity Driven Simulation Development

SEAVEA - Software Environment for Actionable and VVUQ-Evaluated Exascale Applications
UCL - University College London
UNHCR - United Nations High Commissioner for Refugees
VECMA - Verified Exascale Computing for Multiscale Applications
VVUQ - Validation, Verification, and Uncertainty Quantification

LIST OF FIGURES AND TABLES

1. INTRODUCTION

75% of the respondents in a 2021 McKinsey survey of Research and Development leaders said that digital product development was essential to their organisations with advanced product development a key strategic priority. Digital product innovation has become a pivotal aspect of economic growth and social progress with proliferating applications across a diverse range of sectors from aerospace to health, socio-economic development, and humanitarian management among others with $30 trillion dollars in revenue depending on successful product development across sectors over the next five years (McKinsey, 2023; ITU, n.d.; UN, n.d., UNOCHA, 2021). However, emerging technologies like generative artificial intelligence, machine learning, and simulation models etc. find commercial applications after decades of research and development in academic institutions or exploratory projects in big organisations (MIT Technology Review Insights, 2023). Compared to more mature technologies, emerging technologies have longer investment time horizons as they continue to evolve and propel innovation and socio-economic transformation (Chui et al., 2023). However, development of these technology products has to contend with persistent issues in data quality, knowledge gap, contextual awareness, and concerns around risk and responsibility which underlie the enormous investments in advanced digital products and time lag inherent in getting them to wider use (MIT Technology Review Insights, 2023). Despite their rapid development and critical importance across sectors, the innovation processes of emerging digital products and the conditions of their development have received little academic attention.

Within information systems (IS), theorisation on digital product innovation draws from the literature on industrial product innovation and internet studies to posit the layered modular architecture of organising innovation by bringing together a heterogeneous set of actors in an ecosystem (Yoo et al., 2010). A layered modular architecture leverages both design rules of modularity and inherent generative potential of digital technologies to enable proliferating serendipitous development through arms-length relationships with third-party developers (Yoo et al., 2010; Boudreau, 2012; Adner, 2017; Jacobides et al., 2018). A modular design helps decompose a complex design task into its constituent components. These can then be distributed among a wide range of participants through standardised interfaces that can be recombined into novel offerings (Baldwin & Clark, 2000; Schilling, 2000). A layered architecture of the internet consists of four layers: devices, networks, services, and content where each layer represents a different design hierarchy enabling decisions for a given layer to be made independently of others (Benkler, 2006). This facilitates combinatorial innovation across layers which maximises the inherent generative potential of digital technologies (Zittrain, 2006). Premised on paradigmatic examples of mature technological products like mobile operating systems and app development, the layered modular architecture of digital products is said to extend modular design and leverage the unbounded generative potential of digital products through loose coupling among different layers enabled by standardised digital interfaces facilitating distributed development (Yoo, 2013; Yoo et al., 2010; Wareham et al., 2014; Ghazawneh & Henfridsson, 2013).

On the opposite end of a modular architecture lies the integral architecture which involves non-standardised interfaces and tightly coupled components (Ulrich, 1995). While modularity helps in mass production at scale as seen in the examples of proliferating app development, integral architectures are particularly important for highly customised, high-engineering, and high-reliability products like sports cars, airplanes, and high-end electronics. In digital products like robot operating systems which exhibit similar architectural tendencies Lyyra (2018) begins to question the relevance of modular design within what he refers to as complex digital product innovation. This is because the development of complex digital products like robot operating systems have to contend with the requirement of overall specialised knowledge and control of the whole product given its tightly coupled integral dependencies and the distributed nature and open-endedness of digital technologies. The conceptualisation of emerging technologies as complex digital products which are high-cost, high-value, customised multi-technology products involving multiple actors raise interesting questions about their innovation dynamics which cannot be understood through those for mass produced ones (Hobday, 1998; Hobday et al., 2000). Modularity which enables at scale product innovation represents a mature stage of product development where product specifications are fully known in advance and can decomposed through standardisation (Clark, 1985; Salvador, 2007). Given the development journeys of new and emerging technologies, the question arises whether their digital innovation processes can be suitably explained through existing approaches to theorising within the IS field premised on modularity. This is based on the following contention – the theorisation is based on mature technological models with the focus on unbounded generativity and product agnostic scale of the innovation underpinned by mass market conditions. Both of these conditions are absent in the development of new and emerging technologies.

In their study of six large transaction platforms, Fürstenau et al. (2023) highlight the notion of inverse and bounded generativity as components of their extended generativity theory of digital platforms. It highlights two approaches to conceptualising generativity: one as a function of digital product characteristics and two as a function of social interactions within the ecosystem which precludes the digital component. Based on these two approaches, they present two ways of thinking about how generativity plays out in digital product innovation i.e. inverse and bounded generativity. Their conceptualisation rests on two distinct notions of generativity mentioned above. One that is predominantly used in studies of digital product innovation of generativity as flowing from the specific characteristics of digital products and expanding product boundaries through distributed development. The other involves generativity as a mode of work and collaboration among a heterogeneous set of actors. Propositions of inverse and bounded generativity, premised on these conceptualisations appear to present two sides of the same coin. They do not negate the generative potential of digital technologies but point either towards the direction in which generativity of digital product operates or how certain kinds of generativity, conceptualised alternatively as ecosystem boundaries or boundaries of collaboration rather than the product, stabilises. Premised on layered modular architectures, the notion of inverse generativity put forward by them indicates an expansion of product boundaries as a result of user base growth while bounded generativity refers to stabilisation of ecosystem boundaries consequent to growth in the user base. However, using highly formalistic methods like panel vector autoregression analysis, these conditions have been considered separately to understand the influence of one over the other and do not provide insight on the innovation trajectory when the conditions underpinning modularity, such as product maturity and mass market conditions, might be absent. Or when digital characteristics and ecosystem conditions work to mutually shape the innovation trajectory of digital products.

As mentioned earlier, the development of digital technology products proceed through the involvement of a diverse range of heterogeneous third-party actors facilitated by the architecture. Digital technologies have enabled the emergence of flatter non-hierarchical organisational forms like ecosystems which are managed by products owners by leveraging standardised interfaces which connect novel product developments by third parties with the product core (Jacobides et al., 2018; Yoo et al., 2010; Ghazawneh & Henfridsson, 2013). Both the literature on digital product innovation and complex product innovation involve the organisation of multiple actors within the innovation process. Within layered modular architectures, they are organised through standardised digital interfaces that enable a 'thousand flowers bloom' thereby expanding the scale of innovation through unanticipated third-party affiliation that enables the boundless combination and recombination of modular digital components (Boudreau, 2012). Integral architectures of complex products require purposive project-based development by a system integrator who has specialised knowledge of the coherent whole. The innovation process herein is organised through specialised projects involving multi-actor involvement who have to agree on the *ex-ante* path of innovation (Hobday, 1998).

While both strands of literature acknowledge the role of multiple actors, they do not go beyond techno-managerial conditions shaping collective outcomes within the innovation process (Alaimo et al., 2020). Though Tiwana et al. (2010) highlight how innovation trajectories of digital products depend on the coevolution of endogenous design choices and exogeneous environmental dynamics, they only consider the economic environment. Contextualist research within IS highlights the role of a wider set of conditions and socio-technical interdependencies in addition to the immediate context of the IS phenomenon that shape IS innovation (Avgerou, 2000; Avgerou, 2001; Avgerou & McGrath, 2007). However, existing studies within digital product innovation do not adequately explore the wider conditions of possibility within the contexts in and for which such technologies were developed. This highlights a need to explicate the conditions shaping design choices that drives each successive stage of product development. With the emergence and application of newer forms of digital technologies that are applicable in a diversity of contexts, particularly principled and safety critical ones, there is a need to study how digitality in complex digital product architectures shape their development trajectory, the nature of the ecosystems that such product architectures engender, and the contextual dynamics that they are influenced by.

This study uses the case of an agent based simulation model (ABM) for forced displacement being developed by the Migration Modelling and Simulation Group, Brunel University London (BUL) to understand the process of design and development of computational techniques such as ABMs. Forced displacement being a complex multicausal phenomenon and humanitarian management and operations being a globally inter-linked endeavour provides rich variation within its context to study the innovation processes of this digital product. It also holds lessons for theoretical generalisation given the non-commercial development setting. Advanced computational techniques like ABM are increasingly assuming centre-stage in predicting the number of forcibly displaced people to facilitate operational planning for food, shelter, medical care and other relief materials for refugees and internally displaced people (IDPs) (Pham & Luengo-Oroz, 2022). Schelling (1971) first introduced the approach of modelling behavioural patterns of humans, represented as agents, as they interact with their environments based on assumptions and rules. Since then ABMs have diversified in terms of source code language, scale, and model development approaches as computational power continued to expand (Abar et al., 2017). Similar diversification was observed in the areas of its application which ranged from population movements to military scenarios, crowd-mapping, demographic changes, and passenger flow among others. As ABMs enable the translation of small scale behaviour dynamics to large scale system-level outcomes they are finding increasing purchase within population movement studies (Crooks et al., 2008; Castle & Crooks, 2006) like the movement of forcibly displaced people in the event of violent conflict.

The study conceptualises ABM for forced displacement as a complex digital product with multi-technology and multi-actor involvement to understand the nature of interrelationships with the ecosystem that it engenders and the external environmental conditions that determine its innovation trajectory. It draws from the digital product innovation literature in IS that acknowledges interdependencies between the technical architecture and organisational forms like ecosystems as mutually shaping the development trajectories of these technologies (Yoo et al., 2010). This is complemented by context research within IS to recognise the conditions of possibility that shape the interdependencies between the architecture, ecosystem, and environment (Avgerou, 2019). This helps understand the relationship between the technology, proximate context of its development ecosystem and distant context of wider environmental conditions within the design and development of complex digital products like ABM. Acknowledging the iterative development of digital products, the study adapts Kleine’s (2010; 2011) Choice Framework as the approach to contextualise and analyse this relationship and understand how multiple planes of influence across different layers of context shape design choices or outcomes at successive stages of iteration. In order to operationalise this approach it uses process research to identify each iterative phase within the development process and the functional extension in each phase including actors, conditions, and associated choices. This helps develop a narrative account to explain the nature of interdependencies between the digital architecture, ecosystem, and the environment.

The study contributes to the digital product innovation literature through the contextualised study of complex digital product innovation and extends the understanding of digital product innovation driven by such alternative digital architectural forms. This extends our understanding of digital IS beyond modular architectures to take into account emerging technologies like algorithm driven systems. This opens lines of enquiry into the relationship between generativity of digital products, which facilitates recombination and general purpose application, and conditions which shape and structure computational systems and become the building blocks of newer technical architectures in diversifying contexts of application.

1.1 Motivation and scope of the study

As multi-technology and multi-actor socio-technical configurations, complex digital product innovation exhibits multi-level characteristics and dynamics. This highlights the need for contextualised studies of digital product innovation to understand the conditions that shape their innovation dynamics (deReuver et al., 2018). This would help augment understanding of how generativity operates in different technical architectures given the intertwinement between social and technical elements. Understanding the conditions shaping the process of innovation for complex digital products would help in better management of such processes. This becomes particularly important because the development of these technologies will have far reaching consequences as they become the basis of strategic decision-making, resource allocation and socio-economic transformations with the potential to cause disruptive change across sectors (Kallinikos & Hasselbladh, 2009).

This study aims to understand the phenomenon of digital product innovation using the case study of an ABM for forced displacement called Flee. It draws on an existing premise in the literature that the nature of architecture and generativity engendered by digital technologies organise innovation through contribution by heterogeneous actors (Yoo et al., 2010). This study takes a socio-technical view of digital product innovation to understand the social and technical conditions implicated in such a process. The focal phenomenon or unit of analysis involves the digital product innovation process of the Flee ABM model, and this research enquiry explores how it is shaped by the interdependencies between the architecture and its proximate of the ecosystem and distant context of the environment. This responds to the need to understand how these interdependent dynamics play out in understudied digital architectural forms and how its innovation comes to be organised. In doing so it works with *a priori* constructs from the digital product innovation literature such as the relationship between architecture and organisation of innovation, as mentioned earlier in this paragraph. However, many of the key constructs used do not have a general consensus when it comes to terminology, and it becomes important to define the scope and boundaries for this research and what it will include and engage with.

The ubiquity of digital technologies means that the notion of digital innovation has spanned many disciplinary boundaries which has given rise to ambiguity regarding what it actually means or signifies (Hund et al., 2021). Very few research articles within the IS field define digital innovation explicitly, often focusing on an aspect of the phenomenon or conflating digital innovation with its effects. However, extant research on digital innovation can be divided into two broad categories: one that focuses on the development of novel digital products (Yoo et al., 2010) and one that involves consequent changes in existing organisational processes and market offerings of businesses (Nambisan et al., 2017). This speaks to the long standing distinction drawn between product and process innovation (Utterback & Abernathy, 1975). The proliferating use and application of digital technologies have added another form of innovation – business model innovation which involves significantly new ways of creating and capturing business value driven by technology (Fichman et al., 2014). Digital process innovation focuses on the technology adopter where existing ways of doing things, particularly within an organisational setting, stands to change and where such change is embodied or enabled by digital technology. Digital product innovation refers to products that are embodied or enabled by digital technology. Unlike process innovation which focuses on adopters, digital product innovation focuses on the supply side i.e. production, process, institutions, and structures that support and shape the product's development. This study focuses on the digital product innovation and conditions which shape the ongoing development of a digital product. It borrows the notion of complex digital products from nascent research on the same by Lyyra (2018) as a conceptual device for theoretical transferability. The conceptualisation of the Flee ABM model as a complex product helps foreground its characteristics as a highly specialised product which does not exhibit mass market conditions that underpin modular architectures which propel innovation through scale.

Even though the IS literature on digital product innovation is based overarchingly on the notion of modularity and generativity amplified by loose coupling among layers it is driven by the underlying notion that the nature of architecture organises or shapes the social configurations within its proximate context of the ecosystem that drives the innovation process (Yoo et al., 2010; Constantinides et al., 2018; Eaton et al., 2015). The literature on complex products also acknowledge the role of diverse and specialised actors within the product development process (Hobday et al., 2000; Hobday, 1998). However, the notion of ecosystems can have a diversity of meaning within existing literature and therefore it becomes important to clarify the approach used within this study (deReuver et al., 2018). There can be different conceptions of what constitutes an ecosystem ranging from a technical view, a socio-technical view, or an organisational view. Within the technical view ecosystems constitute software based subsystems for the given technical core, the socio-technical view conceptualises ecosystems as involving technical elements and associated organisational or social processes, while the organisational view involves a collection of firms contributing to development of complements (deReuver et al., 2018). This study takes a socio-technical view of ecosystems given distributed development enabled by generativity to understand how architectural forms like ABM organise innovation processes among heterogeneous actors that comprise its proximate context.

Recent studies like Fürstenau et al. (2023) propose an integrative view of generativity that takes into account the ‘product view’ (Yoo et al., 2010; Yoo, 2013; Ghazawneh & Henfridsson, 2013; Zittrain, 2008; Tilson et al. 2010, Parker et al. 2016, Constantinides et al. 2018, de Reuver et al. 2018) and ‘social interaction view’ (Faraj et al., 2011, 2016; Shaikh & Vaast, 2016) focusing on product boundaries and ecosystem boundaries respectively. The ‘product view’ is more closely related to the approach taken in existing studies of digital product innovation to which this study aims to contribute and involves the addition of new categories of functions to an existing product architecture. While its focus might be on the expansion of product boundaries, it also highlights how the particular nature of digital product architectures or digitality help organise innovation through participation of heterogeneous actors (Yoo et al., 2010; Zittrain, 2008; Ghazawneh & Henfridsson, 2013). On the other hand, ‘the social interaction’, focuses on the social interactions between product owners and complement providers which is based on the assumption that product architectures are not per se generative and it is the membership structures that produce innovation through meaningful interaction (Fürstenau et al., 2023). The underlying assumptions between the two approaches tend to conflict since the social interaction view does not attribute a role to the nature of the architecture when, sometimes its specific characteristics like digitality or modularity might make certain kinds of organisational arrangements like ecosystems possible through diverse heterogeneous participation. Keeping in line with the socio-technical approach adopted in this study, while acknowledging the sociality of social interaction view, it is more closely aligned to the product view. This is because while the social interaction view precludes the role of technological artefact, the product view enables to take into account the role of the architecture and digitally enabled generativity in organising its development process among associated stakeholders, thereby simultaneously taking into account both the technical and the social. The product view also helps understand the conditions of ecosystem emergence whereas the social interaction view, in decentring the IS artefact, creates challenges in bringing its development and functional extensions within analytical purview.

Studies within digital product innovation demonstrate that development does not lie on either end of the spectrum between central design hierarchy and a complete lack of coordination and control. It is rather characterised as a tension between distributed development enabled by generativity and control by the product owner (Eaton et al., 2015). Within the scope of modular architectures, this tension is negotiated through standardised digital interfaces (Ghazawneh & Henfridsson, 2013). It involves balancing and resolving the tensions that arise between generativity and evolvability, standard and variety, and flexibility and control through which product offerings and ecosystem coevolve (Eaton et al., 2015, Wareham et al., 2014; Tilson et al., 2010). However, generative systems are also affected by exogeneous conditions that its initial design has not been able to contemplate (Hanseth & Lyytinen, 2010). These shape the progress of digital product innovation and requires attention towards a wider set of conditions to reflect on the dilemmas product development has to negotiate as its progressively tries to incorporate the effects of these conditions (Zittrain, 2008). Jacobides et al. (2018) and Bonina et al. (2021) highlighted the need to understand the role played by the regulatory environment and pressure groups. Zittrain (2006) demonstrated how exogenous conditions like cyberattacks raise a generative dilemma where securing the system might affect its innovation potential. Within the existing literature the exogeneous conditions that comprise external environment have only taken into account economic conditions of competition that shape a given strategy (Tiwana et al., 2010).

As highlighted in the beginning of the chapter, emerging technologies take years to develop a market and might not reflect the modular architecture that establishes the design rules of a development ecosystem that drives digital product innovation through self-selection and self-organisation. Nevertheless, they exhibit unique architectural characteristics, involve a heterogeneous set of stakeholders, and require a better understanding of the wider set of environmental conditions beyond software construction, administrative control, and economic gain (Avgerou & McGrath, 2007). Contextualist research in IS highlights how the social context is deeply intertwined in IS innovation. This becomes even more important in situations where existing economic forces supported by the market are absent. It draws attention to the socio-technical nature of the IS innovation and the need to consider contexts beyond immediate ones (Avgerou, 2000; Avergou, 2001; Avgerou & McGrath, 2007; Avgerou & Madon, 2004; Avgerou, 2019).

1.2 Research objective and approach

This research aims to understand contextualised interdependencies between the architecture, ecosystem, and environment within development of complex digital products like Flee ABM. The principal research question: ***How do technical and contextual interdependencies shape digital product innovation of complex products like the Flee ABM?*** is answered with the help of two operative research questions: *(1)* ***How do complex digital architectures and their development ecosystems coevolve?*** And *(2)* ***How are environmental conditions interrelated with the coevolution of the complex digital architecture and the ecosystem?***

The primary research question aims to understand the role of the architecture, its proximate and distant context within complex digital product innovation. This study aims to unpack the phenomenon of digital product innovation of complex digital products like ABM through an enquiry into interlocking relationships between the generative nature of digital technologies and social and technical elements across different layers of context. The research proceeds to develop explanations for the focal phenomenon from descriptions of the nature of digital product architecture, ecosystem, and environmental conditions. These descriptions form the basis of exploring this trilateral interdependency by first unpacking the relationship between the architecture and its proximate context (first operative research question) and then how the coevolution of the architecture and the ecosystem stands in interrelationship with wider environmental conditions (second operative research question). Using *a priori* constructs from the literature, the first operative research question acknowledges that digital product architectures structure its development ecosystem and is in turn shaped by ecosystem dynamics. This question tries to understand this interdependent coevolution between digital product architectures and the ecosystem i.e. the ways in which complex digital architectures like the ABM structure its development ecosystem and how do the nature of the roles and relationships therein mutually shape its innovation trajectory. Once the first question has explored the dynamic interdependence and coevolution of the architecture and the ecosystem in terms of the proximate context, the second question helps to situate them within wider environmental conditions which expand or limit the development of the digital architecture in particular ways. The two operative research questions help explain the socio-technical interdependencies across layers of context that shape design choices and determine the process of complex digital product innovation.

1.2.1 Theoretical framework

The above discussion highlighted the role of the architecture, generativity, and ecosystem to be the key constituents within digital product innovation with an underappreciated role of the wider environmental conditions. The focus on the relationship between digital architecture, ecosystem, and environment helps unpack the interrelationships between the social and technical elements within digital product innovation to develop holistic theoretical explanations. This helps navigate the under-socialisation and over-socialisation dichotomies that exist within IS theorisation that necessitates adequate explication of the relationship between the technical and the social (Avgerou, 2019; Rush et al., 2021). However, contextualising the relationship between the architecture, ecosystem, and environment requires a delineation of the domains of contextual enquiry. This study considers the role of the proximate and distant context in the form of the ecosystem and wider environmental conditions which shape and structure digital product development through an interdependent relationship with the technical architecture.

The study contextualises this trilateral interdependency through the lens of the Choice Framework (Kleine, 2010; 2011), adapted for this study. The use of Choice Framework helps resolve one of the enduring tensions within contextual research between scale and detail which become particularly contentious within digital products given generativity and multilateral inheritances within ecosystems (Avgerou, 2019; Avgerou, 2013; Winter et al., 2014). The adapted choice framework helps reconcile the tension between scale and scope by identifying design choices at each phase of the development of the product that accounts for reconciliation of conditions of possibility from different planes of influence, thereby helping to explain causal trajectories for digital product innovation. The original framework was developed for the field of ICT4D (Information and Communication Technologies for Development) and is derived from Sen's (1999) work on human development based on the more formal branch of economics called social choice theory which required adaptation for a qualitative socio-technical IS study. This adapted framework draws from research strategies for contextualising an IS phenomenon. It works towards defining and understanding the nature of the Flee ABM architecture, delineating the social collective, and understanding the domain of enquiry to arrive at socio-technical interdependencies between computational design, organisational arrangements, and management and negotiation of environmental conditions. This helps in exploring how the social and technical elements jointly generate outcomes with implications for mutual causality. This also helps understand the multiple interconnected levels of context and reciprocal relationships between context and phenomenon (Pettigrew, 1987; 1990; 1997; Pettigrew et al., 2001).

With complex digital product innovation being a multi-technology, multi-actor phenomenon, contemporary IS development increasingly occurs in multi-disciplinary project based settings (McLeod & Doolin, 2012; Hobday, 1998; Hobday et al., 2000). With different actors coming from multiple contexts, product development comes to be bound by project scope, scale, and complexity which creates multilateral inheritances within the proximate context of the ecosystem (Winter et al., 2014). However, these conditions within the proximate context unfold within the wider socio-economic contexts within which it is located which further structures the opportunities and constraints that the digital product innovation process is presented with. The adapted choice framework helps appreciate the complex and systemic contexts within which such technologies develop. It facilitates understanding and analysis of how particular conditions shape an actor's ability to realise their choices or in other words technological, social and economic conditions which shape outcomes. It rests on the acknowledgment of a resource portfolio which is instrumental in enabling an actor to realise and exercise their agency or choice which are then shaped by environmental factors like policies, regulations, norms, values and other conditions within the external environment. This would help highlight how endogenous design choices which define the trajectory of the digital product innovation phenomenon come to be shaped by the inter-locking conditions within its proximate and distant environment in relation to the digital product architecture. Translation of extant conditions into a given choice within the digital product innovation process depends on the nature of the digital architecture, ownership and control of given resources, nature of interpersonal relationships and conditions of resource provision and distribution within the wider environmental context which mutually shapes the process across multiple planes of influence. This helps in understanding the role of direct and derived reasons which makes the process of arriving at a given choice both deliberative and evolutionary (Sen, 1999).

Application of the adapted choice framework helps in understanding constituent contending tensions within and among different layers of context and technical constraints of the digital architecture. Sen (1999) identifies three ways in which multiple planes of influence within the proximate and the distant environment can negotiated to arrive at given outcomes: by looking at wider set of conditions affecting the phenomenon, partial resolutions where full conciliations between interests and conditions are not possible, and nature of relationships that extend beyond narrow self-interest. This helps to establish causal linkages between the technical architecture, resource portfolio, relationships, and environment which function as conditions of possibility and thereby help explain the innovation trajectory of digital product development.

1.2.2 Methodological approach

In this thesis, the adapted choice framework is implemented through the methodological approach of process research with the Flee ABM as case study. Developing a case study based process research helps unpack the multiple planes of influence implicit in the phenomenon of digital product innovation. This study takes digital product innovation as a process, borrowing from Pettigrew's definition of a process as "a sequence of individual and collective events, actions, and activities unfolding over time in context" (Pettigrew, 1997, p. 2).

Process research within IS is driven by social theory which helps identify linkages between events and actions that contribute to change. This helps establish causal claims that cuts across levels of analysis from the individual to the collective and helps understand how ecosystem dynamics result in aggregate system level outcomes resulting in iterative functional extension and ongoing development of the model. Such multi-level analysis helps connect actions of individual actors in relation to the technology with the wider conditions in which they are embedded (Avgerou, 2013). Following Tiwana et al. (2010), organisation of data into findings starts with a description of the initial state to identify plausible mediating constructs within the development process. Following Pentland (1999) it works towards linking surface level process narratives and descriptions with deep narratives that help explain the sequence of observed events. It involves tracing the sequence of actions and conditions backwards in time to understand observed outcomes and identify events and actions that produced divergent paths and cumulative feedback loops.

In keeping with the adapted choice framework and the above approach for doing process research, analysis commences in a backward direction from outcomes to understand how they are shaped by the interrelationships between digital architecture, ecosystem, and environment and endogenous design choices depending on resources, relationships, and structural conditions (Kleine, 2011). Process research helps understand how outcomes are sustained by complex socio-technical configurations that shape their functionality. The data collected is organised in a chronological timeline of model development that identifies phases, functionalities, stakeholders, and choices associated with each stage of development. Through a historical reflection of the model development process, it identifies intermediate events and actions that result in change from one state to another acknowledging that each state of change can be driven by multiple influences that translate individual events or action to collective outcomes (Devaraj & Kohli, 2003). While process narratives aim to establish causal linkages between events, actions, and outcomes they are distinct from uni-directional cause-effect relationships in variance models. They admit circular and reciprocating models of causality with multiple causal chains (Orlikowski & Baroudi, 1991; Langley, 1999). Further, while variance models are suited for theory testing studies, theory testing and theory building cannot be separated within process narratives where explanations are constructed to establish the causal logic behind a given phenomenon (Avgerou, 2013).

1.3 Targets for contribution

The research aims to contribute to the literature on digital product innovation in two ways: Firstly, developing a theoretical and conceptual approach to include emerging digital technology architectures like algorithmic models within theorisation on digital product innovation. Secondly, to propose a contextualised process-based approach for studying digital product innovation that takes into account the diversity of conditions that shape the innovation phenomenon of digital products. It uses the notion of complex digital products as a conceptual device for generalising the output of this research towards incremental contribution in thinking about new computational architectures and their innovation dynamics. While the 2022 debut of ChatGPT might herald a new era for emerging technologies finding proliferating application, significant amount of advanced technologies are still in experimental phases within research and development departments and academic institutions. A better understanding of their innovation trajectories and the conditions that structure them will help better manage their development process. Moreover, appreciation and acknowledgement of the conditions that shape design choices will help manage potential social implications which flow from them (Markus & Nan, 2020). While the study proceeds from *a priori* constructs within the literature on interdependency between architecture and the ecosystem (Yoo et al., 2010) alongside the role played by wider environmental conditions (Avgerou, 2000; 2001; Avgerou & McGrath, 2007), this relationship is explored and unpacked by understanding the nature of its constituent parts. The nature of the architecture, ecosystem, and environment are explicated in order to arrive at a characterisation of their relationship. These descriptive endeavours become important explanatory devices to understand how specificities in each of these elements shape the conditions for further development and innovation. This becomes particularly important since technical architectures beyond modular ones have not received adequate attention within the digital product innovation literature and therefore the digital innovation dynamics that pertain to them have remained neglected. This highlights the need to also study the context within which such development takes place since initial development proceeds from search and learning to arrive at a contextual fit (Tilson et al., 2010; Clark, 1985). Understanding how different layers of context have implications for the product development process highlights conditions that would need to be factored in for their management. The deployment of process research and adapted choice framework helps understand the interplay between digital architectures and contextual conditions that shape the process of digital product innovation. Working to understand the conditions that shape the successive phases of development helps unpack relationship and environmental conditions on which innovation trajectory is contingent. It helps resolve the tension between how generativity is negotiated within the interdependent relationship between alternative digital architectures, their development ecosystem, and the wider environment (Tilson et al., 2010; Wareham et al., 2014; Eaton et al., 2015).

To arrive at these intended contributions, the study first organises the data through process approach in terms of the phases and resources, relationships, structural conditions, at each phase. These phases are then used to first develop a process narrative to provide an overview of the development process before being disaggregated into constituent elements to describe the nature of such constituent elements. This helps understand the process of incremental development of the code and software for Flee, the ecosystem dynamics that bring them into being, and the environmental conditions that structure them.

1.4 Structure of the thesis

The thesis is structured as follows:

Chapter 1, Introduction: Provides an overview of the thesis including motivation, scope of research and targets for contribution. This helps bound the research effort within a well-defined scope of investigation and contribution by delineating the phenomenon and explanatory conditions to be explored.

Chapter 2, Literature review: Explores the literature on digital product innovation spanning product architectures and ecosystems as defined within the scope in Chapter 1. It also involves a review of the particular nature of ABM to highlight its complex product characteristics of being a multi-technology and multi-actor phenomenon which helps in understanding the nature of the technology under study. It highlights the importance of wider conditions of possibility and of a contextualised approach that helps negotiate attendant tensions within literature. Positioning within the literature helps scope out the research questions that are explored in this study.

Chapter 3, Theoretical framework: Presents the adapted choice framework through synthesis of approaches to contextual research in IS and the Choice Framework developed by Kleine (2010; 2011). The adaptation process involved rationalising aspects of the original framework complementing a contextualised approach in relation to *a priori* constructs.

Chapter 4, Methodological approach: Highlights the case study based process research approach and how it complements and helps mobilise the adapted choice framework with an analytical approach guided by *a priori* constructs derived from literature and theory. The chapter outlines the rationale for the methodological approach and case selection, the process of organising the data, the pathway for making a contribution.

Chapter 5, Findings: This chapter organises the findings from the year-long non-participant observation, collation of research articles, internal documents, and preliminary interviews that helped set the scene for the observation process. This demonstrates a process research mode of organising the data. Appendices 1 and 2 contain a tabulation of phase wise progression of the Flee model which identifies the resource portfolios, ecosystem participants and environmental conditions at each phase. The consolidation of findings provide a descriptive account of the nature of ABM architecture, ecosystem dynamics, and environmental conditions while also identifying mediating conditions of resources, relationships, and structural conditions which serve as building blocks for analysis.

Chapter 6, Analysis: Proposes a conceptualisation about the process of digital product innovation by explaining the nature the of relationships between the architecture, ecosystem, and environment drawing on descriptive accounts and mediating conditions developed and identified in the previous chapter.

Chapter 7, Discussion and conclusion: Reflects on the theoretical contributions and limitations of the thesis with implications for practice and future research.

2. LITERATURE REVIEW

Chapter summary: *This chapter positions itself within the strands of literature on which existing theorisation on digital product innovation is built while exploring their underlying assumptions and relevance to emerging digital products. This study contends that the digital innovation of complex products like ABM represents different trajectories than ones predicated on layered modularity which has been studied in the literature so far. The coevolution of complex digital product architectures and their associated development ecosystem operates under conditions of high uncertainty and through purposive integration of a diverse range of knowledge, resources, and components. As a result, it comes to depend not just on its proximate context of the ecosystem but also needs to respond to wider environmental conditions. Therefore, it becomes important to explicate these conditions and socio-technical interdependencies that shape design choices at each iterative stage that leads to successive functional extension of the product.*

2.1 Digital product innovation

The critical aspect of digital product innovation is digitisation which lends digital products their specific character of homogeneity, reprogrammability, and self-referentiality which means that digital artefacts are perpetually in the making (Yoo et al., 2010; Kallinikos et al., 2013). This enduring incompleteness means that the nature of tasks and operational links that a digital product might or can accommodate is not foreclosed resulting in unbounded generativity (Zittrain, 2008). The unbounded generative nature of digital products, in general, pervades through and is facilitated by the product architectures (Wareham et al., 2014). In this way product architectural design sets out the structural composition that determines at one level the overall functionality of the product but also how the tasks for its ongoing development can be distributed among a heterogeneous set of actors i.e. how its innovation processes comes to be organised.

Theory development within digital product innovation has drawn on modular product architectures from industrial innovation and layered architectures of computer systems to posit the layered modular architecture (Yoo et al., 2010). In such architectures, modularity decomposes the design problem into a stable core and complementary modules. This determines the design rules of task distribution among heterogeneous third-party developers

(Baldwin & Clark, 2000) while generativity through loose coupling among layers results in proliferating novel offerings resulting from combination and recombination of digital complements (Benkler, 2006; Zittrain, 2006; Wareham et al., 2014). Digital complements are more versatile and product agnostic as opposed to traditional product complements and have high degrees of recombinability and indeterminate functionality (Kallinikos et al., 2013). Further, generativity of systems which subtend digital product innovation enable recursive processes of development that supports high degrees of distributedness by providing a space for 'revisable configurations' (Alaimo et al., 2020, p. 28). This enables multiple stakeholders to converge by contributing their knowledge and skills towards the development of a digital product (Constantinides et al., 2018).

Digital products do not represent linear value chain models of product development but fit an ecosystem model where product development is shaped by participants and stakeholders involved (Jacobides et al., 2018). Traditional product innovation has been predicated on modularity which enables the distribution of design tasks among a wide group of people based on design hierarchies and rules managed by a central design agency (Baldwin & Clark, 2000). While the literature has acknowledged the role of modularity as the key driving design principle which privileges the emergence of an ecosystem (Jacobides et al., 2018), the development of complex products like ABM also require the management of diverse external relationships as they involve the integration of multiple technologies, components, and resources (Hobday et al., 2003; Hobday, 1998; Hobday et al., 2000). Further, modular design is facilitated in the presence of mass market conditions where design rules through standardised interfaces facilitate serendipitous development through third-party affiliation. However, complex product development takes place in the absence of such conditions and needs to structure around purposive component integration and knowledge and resource assimilation which requires an alignment of activities, actors, position, and links (Adner, 2017; Hobday et al., 2003).

Digitality and associated unbounded generative potential appear to challenge the notion of a centralised design agency through its capacity to facilitate distributed development through a layered modular architecture (Yoo, 2013; Yoo et al., 2010). However, studies have shown that the product owner still retains substantial control and modulates this control to balance competing objectives and intended outcomes (Ghazawneh & Henfridsson, 2013; Tilson et al., 2010). However, these modes of control are not uncontested and 'wakes' of influence within digital ecosystems shape aggregate outcomes through extant power differentials among stakeholders (Eaton et al., 2015, p. 219). Digital product development proceeds through negotiating the tension between generativity and control (Yoo et al., 2010; Eaton et al., 2015; Ghazawneh & Henfridsson, 2013; Tilson et al., 2010) in order to ensure that they remain both 'stable and evolvable' (Wareham et al., 2014, p. 1196). These endogenous design choices of product owners are shaped as much by internal conditions as by the dynamics of their external environment (Tiwana et al., 2010). However, these studies have been premised on uncovering managerial and economic rationales and incentive structures (Sabamurthy & Zmud, 2000) which obfuscate the way a particular environment and actor configuration coincide with technical architectures and systems (Alaimo et al., 2020). Within modular architectures, design rules based on standardised interfaces determine partition of decision rights that help product owners shape innovation trajectories towards desired outcomes. However, they come to depend on the incentive structure based on comparative power differentials within the ecosystem (Ghazawneh & Henfridsson, 2013; Eaton et al., 2015). Incentive structures can differ when product innovation takes places in the absence of mass market conditions which favour modularisation (Utterback & Abernathy 1975; Clark 1985; Miller et al., 1995) and in non-business contexts shaped by a wider set of conditions of possibility that could result in a given outcome (Avgerou, 2019).

2.1.1 Products and product architectures: Underlying conditions and characteristics

The nature of product architectures determine how their innovation trajectories are organised (Yoo et al., 2010). Product architectures provide the 'conceptual blueprint' whereby functionalities of the product are assigned to its component parts (Ulrich, 1995). It involves the arrangement and mapping of functional elements and the specification of interfaces amongst its interacting components. Functional elements represent abstract and conceptual responses to requirements shaped by the expected functioning of the product under given constraints. This is because every design problem tries to find fitness with form and context where the context defines the problem and the form is the solution to the problem so defined (Clark, 1985; Alexander, 1964). The combination of functional elements and dependencies among them shape the functional structure that defines how the product will operate. The arrangement and interdependencies among functional components that make up a product architecture determine whether they can be described as modular or integral (Ulrich, 1995). As the different functionalities are distributed over a range of components that collectively contribute to the overall functionality of the product, they may work in one of the following ways: (1) individual component performing a single function; (2) individual components are a part of an overall composition of components where each contribute to the production of a single functionality; (3) perform multiple functions simultaneously on their own (Ulrich, 1995). A product architecture is modular if the change in functional elements and components can be decoupled in a way that change in one component does not require a change in the components that it is connected to. Product architectures are integral if a single component implements multiple functional elements and components are coupled in a way that change in one necessitates a change in another (Ulrich & Eppinger 2012). These specifications come to depend on the way functionality has been allocated during architectural design.

Modularity represents a design principle that decomposes the overall design requirements and functional structure into core and accompanying modules that can be combined to extend the functionality of the core and the overall product (Baldwin & Clark, 2000). Modules have structural connections to the core through standardised interfaces within the boundaries of a module but such connections are loosely coupled (Schilling 2000; Parnas, 1972). This loose coupling helps maintain the overall common functionality while allowing combinability of components in furthering product development (Salvador, 2007). The overall architectural system comes to define design parameters and rules that help partition design tasks on the basis of combinability and separability among a dispersed group of people. This gives rise to design and task structures, activities, and economic systems that mirror the architecture of the product (Baldwin & Clark, 2000). It embeds technical and organisational coordination mechanisms to enable distribution and coordination of tasks among a group of people (Sanchez & Mahoney, 1996). Thus, modularisation, through loose coupling, enables the task partitioning of the overall design for it to be distributed among diverse groups of individuals. This helps in maximising variation in products and enables participation from a diverse developer pool while keeping costs low through combination of components in different configurations. The strategic importance of cost-effective product variation is a function of mass-market conditions which drives modularity in product architectures. Mass market conditions also allow dominant designs to emerge and design hierarchies to be controlled (Utterback & Abernathy 1975; Clark 1985). The control of design hierarchies requires the presence of a centralised agency which is able to establish and maintain design rules and use them to coordinate among stakeholders. Modularisation becomes possible when the overall functional requirements are known in advance (Salvador, 2007). Consequently, they represent the standardised stage of product evolution unlike initial stages where requirements and functional criteria might not be fully defined or available (Clark, 1985). Thus, modularisation represents the end stage of a product innovation process where dominant designs can lead to increase in production volumes as functional criteria comes to be more fully defined. As a result, modularity becomes a function of scale by managing the complexity of a design problem by distributing tasks among a large group of people when it becomes difficult for a single entity or organisation to bring about at-scale product development alone (Garud & Kumaraswamy, 1995; Langlois & Robertson, 1992).

Complex products represent an analytical category for highly customised technology intensive products that preclude modularisation in the absence of mass market conditions. As a result, they present innovation dynamics that differ from mass produced products (Hobday, 1998; Hobday et al., 2000). This is because (a) they are comprised of many customised and interconnected elements designed specifically for a particular user or given application scenario; (b) they are continually evolving as small changes in underlying technologies or certain sub-systems can require significant alteration in other parts; (c) Their development is organised in projects or small batches that allow high degree of user involvement in the innovation process rather than arms-length market transactions that are common in mass produced goods (Hobday et al., 2000). Complex products often exhibit integral architectures exhibiting tight coupling and overlapping mapping across different functional elements (Ulrich, 1995). They represent high-performance, high-reliability, high-value, and cost-intensive products like telecommunication exchanges, aeroplanes, intelligent buildings etc. (Hobday et al., 2000). The term 'complex' is used to reflect the high degree of customisation, the breadth of specialised knowledge and skills required in their production, as well as critical product dimensions. Complex products require several producers to work together simultaneously and is a multi-technology endeavour when production techniques move from mass production to unit production involving the integration of different technologies and extensive collaboration among different stakeholder groups. Development of complex products involve a system integrator that works in conjunction with other stakeholder groups like users, suppliers, and regulators who organise in successive temporary multi-actor alliances in the form of specific projects to solve a particular design problem. The projects act as focusing devices that enables particular problems, so identified, to be addressed in detail. This is because the development of complex products operate under considerable uncertainty due to changing user requirements, and unexpected or uncertain events and interactions, changes in relevant policies or regulations, or developments in underlying technologies. As a result, production proceeds through instances of purposive temporary collaborations like project-based coalitions of relevant stakeholder groups where different aspects of the product development process involves *ex ante* negotiation among relevant actors.

This section reflected on two types of product architectures, how they are structured, and the conditions that enable them. IS research on digital product innovation draws from the strand of literature on industrial innovation premised on modularity to understand how digitality, or the nature of digital technologies, shape these processes and create newer conditions for digital product development. Modularisation has played a key role in application software development through task-partitioning and distributed development. However, in the context of emerging digital technologies like robots and autonomous systems, Lyrra (2018) demonstrates how innovation trajectories of such complex products come to depend on integral architectures that involve low-level hardware control to high-level software that performs decision-making, reasoning, and planning tasks forming multiple aspects of managing the design of a digital product. The following section looks at the role of digitality and product architectures as studied in the literature.

2.1.2 Digitality and product architectures

Pervasive digitisation has led to a rethinking of traditional product architectures and how digitality comes to be implicated within ongoing product development (Yoo et al., 2010; Henfridsson et al., 2014). Digitality has challenged dominant designs as a result of their characteristics of granularity, recombinability, and data homogenisation which enable unbounded generative potential of digital technologies (Tilson et al., 2010; Kallinikos et al., 2013; Yoo, 2013). This stands in contrast to single design hierarchies, fixed interfaces and design features of final products that recur across an industry within traditional product innovation. This is because digital technologies can be programmed to circumscribe a wide range of functions thereby making them product agnostic (Kallinikos, 2006; Yoo et al., 2010). Digital products are perpetually in the making due to their generative properties (Garud et al., 2008; Zittrain, 2008). Generativity leads to unbounded potential of creating new products from existing products or components. It is underpinned by the characteristics of digital objects and their ability to be recombined and reproduced at low marginal cost (Faulkner & Runde, 2009; 2019). Pervasive digitisation i.e. rendering reality into binary digits leads to data homogenisation making digital products reprogrammable, self-referential entities. These digitised bitstrings carry the enduring properties of digitality that shape the nature of digital products with fluid boundaries and unbounded generative potential. These bitstrings follow a set of syntactic rules, are structured through distinct organised parts, where every part endures simultaneously once created (Faulkner & Runde, 2019). These aspects of digitality underpin the generative potential of digital products supporting ongoing innovation through combination and recombination.

Drawing from his study of the internet architecture, Zittrain (2008) rests the notion of generativity on the arrangement of digital networks and artefacts that are layered as stacks which can be connected to each other through standards and interfaces that act as gateways between layers. This lowers the threshold of participation by enabling particular types of activity across its different layers. He defines generativity as the 'system's capacity to produce unanticipated change through unfiltered contributions from broad and varied audiences' (p. 70). It involves pairing an unfiltered input received from a participant with the output of 'unanticipated change' (p. 70). Individual generative tools can be organised into generative systems or larger technological arrangements developed among a larger group of people; generative systems then become the foundation from which digital product innovation emerges by bringing about wider and unfettered participation.

Yoo et al. (2010) combines the layered architecture of the internet with the modular architecture of product systems to outline the layered modular architecture of the digital products. The layered modular architecture operates at different levels i.e. of the device, operating system, and applications. As digitality become implicated within a layered modular architecture it decouples services from devices and content from network thereby opening up possibilities of outside innovation. A digital modular architecture involves an extensible codebase with functionally independent complementary modules (deReuver et al., 2018). The codebase provides core functionality shared by the modules that interoperate with it through standardised interfaces. These modules can be seen as "add-on software sub-systems" (Tiwana et al., 2010, p. 676) or "executable pieces of software that are offered as applications, services or systems to end-users" that extend the functionality of the core product (Ghazawneh and Henfridsson, 2013, p. 175). The loose coupling between its layers enable product agnostic development where generativity of digital components and their recombinability lead to new product configurations enabled through participation of a distributed range of actors. Unlike traditional modular architectures, digitality leads to multiple inheritances in distributed settings, thereby upsetting the traditional single source of control over the product development. Its facilitation of distributed development through digital modules, which are product agnostic and indeterminate, essentially postpones a decision on design features through a late binding of capabilities by third-party developers (Svahn & Henfridsson, 2012).

The layered modular architecture provides an important way of thinking about mass market digital products with well-established design rules like mobile application development (Yoo et al., 2010). However, developers in early stages of complex product development often have to contend with developing a product design that is able to meet the requirements of initial users while trying to ensure the completeness of their designs (Hanseth & Lyytinen, 2010). This process comes to involve ‘discovery, implementation, integration, control and coordination of increasingly heterogeneous IT capabilities’ and actors with diverging interests (p. 2). A generative system facilitates product development through stakeholder contribution as a function of both technological design and social conditions and the way a system relates to its members and the members to each other (Zittrain, 2006). Moreover, computational and predictive technologies are not agnostic to the context of their application and involve an analytical reduction of reality and their reconstitution in the form of predictive output (Kallinikos, 2009). The computational rendition of reality involves the delineation of the operational domain and a simplification of causal parameters within a phenomenon of interest.

Lyyra (2018) conceptualises the development of complex digital products like robots and autonomous systems as contextually bound and embodied chains of transformation as developers need to iteratively integrate the variability of the physical environment that a robotic system needs to navigate on implementation. Similarly, complex products like ABM for forced displacement present different considerations. Highlighting these specificities helps to understand considerations for different allocation schemes within the functional structure of the product. As a result, the final design can exhibit the desired functionality within the eventual architecture resulting from design choices that are determined by a range of factors spanning functional requirements to production systems and strategic direction (Campagnolo & Camuffo, 2009). Thus, digitality operates through product architectures on the basis of the sociotechnical input i.e. digitised bits of social reality that are shaped by their contexts, involvement of internal and external developers, with inherently unbounded properties (Hund et al., 2021).

As discussed earlier, Fürsentau et al. (2023) propose an extended generativity theory that aims to combine, rather than synthesise, two approaches to generativity which they refer to as the ‘product view’ (expansion of product boundaries) and the ‘social interaction view’ (expansion of ecosystem boundaries) to explore their relationship with user base growth within layered modular architectures of digital transaction platforms. They explore the role of user base growth on product boundaries and ecosystem boundaries separately to arrive at the notion of bounded generativity (stabilisation of ecosystem boundaries) and inverse generativity (expansion of product boundaries). The notion of user base growth is linked inextricably to a layered modular architecture. Consequently, the conceptualisations of inverse and stabilising relationships, can potentially be explained through conditions of ecosystem evolution in existing literature. Within modular architectures the dynamic growth of the ecosystem can potentially stabilise over time, which the authors characterise as bounded generativity, on account of market maturity or market saturation which leads to stabilisation of demand in addition to the authors’ acknowledgement of platform maturity and resolution of conflicts around key components over time. These can also be the result of technological convergences and envelopment due to overlapping users and developers for different layered modular products as has been acknowledged by Eisenmann et al. (2006) and Tiwana et al. (2010). However, these factors have not been accounted for in the existing study (Fürstenau et al, 2023). Further, it can be argued that *effects* of generativity, rather than the nature of generativity itself, explored in this paper with regard to the notion of inverse generativity can be tied back to the indeterminate product-agnostic nature of digital products which underlie its generative potential whereby product development and evolution needs to iteratively capture the evolving needs of an expanding user base based on search and learning (Yoo et al., 2010; Kallinikos et al., 2013; Hanseth & Lyytinen, 2010; Evans et al., 2006; Clark, 1985; Kling, 1992; Williams & Pollock, 2008). Moreover, complex products tend to involve institutional users, since they are not a function of mass market conditions, who have high levels of involvement in the development process. Further, ecosystems around complex products tend to proceed through purposive integration as opposed to growth by affiliation so it remains to be seen how the inherent generativity of digital product shapes and is shaped by such relationships which also need to respond to wider environmental conditions in which they are situated.

2.1.3 Coevolution of architecture and ecosystem

Previous sections acknowledge the role of a diverse range of actors in the product development process where the nature of such participation is determined by the architecture of the product. The literature within IS on digital products acknowledges the interdependency between the architecture and the emergence of flatter organisational forms like ecosystems (Jacobides et al., 2018). This is because digital system development has moved beyond maximising task efficiency in standalone systems and no organisation has the resources, power, or legitimacy to understand such development and produce change alone (Tiwana et al., 2010; Van de Ven, 2005). The interdependency between technical and organisation form depends on the degree to which form and function are able to be disaggregated into subsystems (Simon, 1962), the degree to which changes in subsystems shape the overall functionality of the product, and the design rules by which constituent development in organised (Tiwana et al., 2010).

Product development then comes to depend on technical characteristics of the architecture and dynamics within its development ecosystem. Product architectures set up the conditions through which coordination is enabled within an ecosystem. The overall dynamics of the ecosystem and its direction of evolution are determined through decision-rights partitioning or how decision-making authority is divided between the product owners and other stakeholders within the ecosystem. Control mechanisms can involve output control where each output of constituent development is evaluated by the product owner, process control where the product owners specifies procedures for other stakeholders to follow, or informal control in the form of norms and beliefs. Control mechanism are used to facilitate coordination rather than manage widely divergent zero-sum interests (Tiwana et al., 2010). Moreover, while the product owner or the ecosystem architect wields substantial power in determining the direction of innovation (Ghazawneh & Henfridsson, 2013), there are multiple planes of influence which makes the nature of control bidirectional and even multidirectional (Eaton et al., 2015).

Existing research within IS highlights the role of modularity as being one of the preconditions behind the emergence of ecosystems where technological modularity allows different parts of a system to be produced by different producers (Baldwin & Clark, 2000). While the overall design parameters are set by the product owner, actors or entities producing complementary modules have a degree of autonomy over their unit of production (Jacobides et al., 2018). The technical architectures and the organising principles jointly determine future development and conditions of digital product innovation. Consequent product development then involves maintaining a balance of the control and autonomy by the product owner that enables it to direct ecosystem activities while encouraging development among its affiliate developer community.

The tension between control and innovation can be represented by the processes of resourcing and securing managed by product owners using digital interfaces within layered modular architectures (Ghazawneh & Henfridsson, 2013). Resourcing refers to the process by which the scope and diversity of the product is increased whereas securing refers to the process of increasing control over a product and its offerings. Product owners maintain control by securing within the ecosystem through administrative actions, contracts or strategies that enable the owner to maintain its leadership position in attracting third-party development. Diversity within an ecosystem is enabled either by third-party developers pushing product boundaries when they find the existing ones to be inadequate or transforming them in a way that stimulates new application areas. However, Eaton et al. (2015) contend that power differentials play an important role in the use of digital interfaces for coordinating ecosystem outcomes. Heterogenous actors associated with the product development bring about change through ‘cascading wakes of influence’ (p. 219) which results in different degrees of control over resources and other actors.

Within digital ecosystems data is said to be a key resource that mediates digitality and subsequent conditions for collaboration and coordination (Alaimo et al., 2020). Data often becomes the fulcrum that underpins operational and functional linkages within an ecosystem. Data and associated functionalities become instrumental within architectural design in designating the role of participants within the ecosystem. Successive product development and functional extension within digital ecosystems come to depend on how different data types and formats are produced, combined, and used. This involves the standardisation and formatting of data to enable functional combination with other data and digital systems which determine the nature of complementary relationships and mutual value while remaining reconfigurable and updatable (Kallinikos, 2006; Marton et al., 2013). Such product development comes to be contingent on how such data and functionalities come to be linked to the roles taken by different participants in the ecosystem with end-users often switching between producers, reviewers, and consumers. While digital product development is driven by ecosystem evolution based on mutually reinforcing complementarities such complementarities depend on the systematic exploitation of resources like data which are combined into more complex functionalities of the product. Using TripAdvisor as the empirical example, Alaimo et al. (2020) discuss how different types of data are assembled to construct popularity indexes, booking packages, and data analytics subscriptions among others. These complementarities are driven not just by the mutual evaluation of comparative interests but also by the linkages between inputs, functionalities, and technologies and the systems subtending them. The standardisation and computation of collected data are aggregated and mashed up across the value chain (Alaimo & Kallinikos, 2017; 2019) and much depends on the availabilities of such data resources and the conditions of such availability (Alaimo et al., 2020). Thus, product evolution and functional extension come to depend on resources like data that came into play at each stage.

Within the TripAdvisor study, the first transition was from a travel search engine to a social media travel website which involved a shift from the use of traditional and already available data to new sources like user-generated data in the form of ratings and reviews. The second transition from social media travel website to provision of end-to-end services from search to booking coincided with the use of transaction data that serve multifunctional operational requirements from booking to providing rankings based on analytics on user-generated content. This is done by combining different types of data into an aggregated data pool from which different metrics and scores are computed (Alaimo & Kallinikos, 2017; 2019). However, this uptake of data into product development is underpinned by technological infrastructures and systems that subtend them. The first stage of the evolution went hand in hand with an underlying technological switch from search and indexing to Web 2.0 characterised by the use of social and interactive technologies. The second transition involved the development and implementation of technological systems that supported advanced analytics, price comparisons, and ranking by mashing up different types of data. These transitions were predicated on the uptake of different types of data and the evolution of underlying technological systems linked to the configurations of actors, roles, and positions at different stages of product evolution (Alaimo et al., 2020).

Modular architectures encourage the growth of ecosystem by affiliation where multiple interests need to be managed (Darking et al., 2008). This has been viewed as a dialogic relationship between stakeholders (Wareham et al., 2014). However, digitality and associated generativity raised paradoxical considerations of change and control (Tilson et al., 2010). Within an ecosystem as affiliation this relates to the need for stability in order to encourage further growth and while maintaining flexibility to encourage complementary development through affiliation. This necessitates a mode of both centralised and distributed control that must guide ecosystem activities in desired directions without compromising on its unbounded generative capacity. Within IS, ecosystems have been studied in the form of ecosystem as affiliation that enables the wider participation through standardised interfaces. More widely too, the study of ecosystems has mostly been predicated on modular architectures because of their decomposable nature that reduces entry and participation costs (Jacobides et al., 2018). However, as discussed earlier, in the case of highly specialised projects for complex product innovation, systems integration needs to be organised around a focal product innovation.

Adner (2017) suggests ecosystems can also be viewed as a structure which contrasts with the view of ecosystems as a collection of focal and affiliate organisations. Within the view of ecosystem as structure, the organisational form develops around a focal innovation as opposed to standardised interfaces defined by top-down design rules. While the role of a system integrator is not minimal it involves substantial up and downstream dependencies and linkages that allows the product to materialise. This involves a structure of complementary roles, relationships, and activities between stakeholders can cannot be reduced to the sum of bilateral relationships (Jacobides et al., 2018; Adner, 2017). Such complementarities need to be non-generic in nature that require the need for coordination for their integration within the system.

As digital products and ecosystems coevolve, consequent development comes to depend upon the internal fit between architecture and ecosystem activities and the external fit between endogenous choice around architecture and ecosystem and the dynamics of their external environment (Tiwana et al., 2010). Choices that might be appropriate for one environmental context might be inappropriate for another. Internal choices of product owners shape expectation and facilitate coordination within the ecosystem in order to achieve compatibility with its environment (Katz & Shapiro, 1994). Misfits between internal choices and the context determine the survivability of the product and thereby the subsequent product development process. Endogenous choices regarding the technical architecture would need to accommodate the conditions and possibilities for product evolution to respond to demands and requirements from a widening user base and should be able to support variety and evolvability over time (Hanseth & Lyytinen, 2010). This is further said to be shaped by exogeneous environmental dynamics that involve the speed of technological evolution and their integration with application domains adjacent to the ecosystem (Alaimo et al., 2020).

As a complex socio-technical undertaking digital product development is sustained by a range of technologies and technology mediated operations held together by technical links and architectures. These come to condition the limits of their functional extension at a given point in time due to the need to manage both technical constraints and diverse interests. Some of the background technological conditions and complex technical arrangements that subtend a particular product shape the potential for action and innovation while providing a space for 'revisable configurations' that is able to respond to the evolving and dynamic needs of an expanding user base (Alaimo et al., 2020, p. 28). Ecosystem dynamics are shaped around operational and economic advantages among participants due to value reinforcing complementarities that extend beyond bilateral relations (Adner, 2017; Jacobides et al., 2018). However, as the combination of resources brought about by these dynamics acquire greater value these are increasingly shaped by the digitality of such resources (Alaimo et al., 2020). Within extant IS literature, product development ecosystems are limited to an 'architecture of intentions' driven by economic considerations but fail to deal with the means by which they materialise into actual relations (Alaimo et al., 2020; p. 43). Digital resources shapes the 'structure of means' that over time integrate with the final product to circumscribe possible actions while excluding others (Alaimo et al., 2020; p. 43).

Thus, product development is shaped by the nature of the architecture and conditions of the ecosystems in which they are embedded and by the resources and links that shape the nature of relationships among them. The structure of links and interrelationships are further shaped by the wider technological systems and architectures, that underpin their operations (Yoo et al., 2010; Alaimo et al., 2020). Understanding this relationship structure within an ecosystem involves an understanding of the nature of stakeholder involvement and the critical roles they play in conjunction with the technology in a web of functional relationships (Kallinikos et al., 2013). Ecosystem dynamics depend on the synergies and complementarities between activities, resources, and output within a multilateral arrangement of stakeholders (Alaimo et al., 2020). These synergies or complementarities are reinforced on the basis of the incentive structures that maximise value for participants within an ecosystem (Adner, 2017; Jacobides et al., 2018; Teece, 2018). Each stage of product development coincides with a particular environment, actor configurations, and interdependencies (Alaimo et al., 2020), even more so for project-based organisation of innovation for complex products. Studying these conditions with purely economic and managerial considerations tend to obfuscate the way such conditions work in conjunction with technological architectures and systems. Over time the technological architecture comes to establish the framework within the which certain actions and choices develop and require careful consideration. Linking technology to action requires approaching technology as a structuring force that shapes social and economic relationships (Faulkner & Runde, 2013; Kallinikos et al., 2013).

2.2 Complex product characteristics of ABM for forced displacement

Simulation models have played an increasing role in decision-making and policy development, particularly during the COVID-19 pandemic (Leonardi et al., 2021). They provide a way of modelling complex systems in the face of uncertainty and information overload. Simulation models can be conceptualised as complex products as their development involves composite combinations of data, mathematics, and assumptions involving approximate contextual and behavioural factors. Given their highly specialised applications, they are developed in the absence of mass-market conditions and require high levels of technical and system awareness on the part of the integrator to bring together diverse and specialised expertise on components, knowledge, and resources involved in the final product (Hobday et al., 2003). The development of simulation models are based on assumptions about dynamics that govern complex systems which can be derived from theory, past data, or both (Leonardi et al., 2021). Modelling a particular phenomenon can involve widely varying assumptions in defining the functional area, selection of parameters, and validating its outcomes. Moreover, a given phenomenon can be approximated using multiple models with competing assumptions and different data sources to explore the boundaries and variations around a given issue. Modellers sometimes use a weighted average of the models they have built which has a compounding effect as a result of including multiple overarching models. This leads to an ensemble architecture that represents an average of averages (Leonardi et al., 2021).

ABM is a type of simulation modelling that enables the aggregation of individual behaviour patterns into system level outcomes. It is a method for modelling heterogeneity in individual decision-making (Bonabeau, 2002). It can explicitly model social interactions, resulting networks and emerging behaviour at higher levels. ABM can be used to model active objects or agents in relation to time, event, or behaviour (Borshchev & Filippov, 2004). Elements within an ABM involve agents, their environment, and the relationship among them (Macal & North, 2014). Agents are autonomous, unique, and distinct in their attributes, behaviour, size, and location where the use of ABM enables demonstrating how agents and their environments vary across time and space. Agents can have static (autonomous; self-contained; social) and dynamic attributes (explicit goals; ability to learn and adapt) and their interaction can be dependent on individual behaviours, behaviours that influence other behaviours, or rules that determine dynamic attributes.

While the concept of ABM is as old as the 1940s it only came into its own in the 1990s because its development and use required computational advancements (Arora et al., 2017). While ABM has found diverse areas of application in infrastructure, military, cyber-security, and climate change it is increasingly being applied in areas of population movement (Allan, 2010; Castle & Crooks, 2006; Crooks et al., 2008). While the conception of representing people as agents and modelling their social behaviours as agent interactions was first proposed by Schelling (1971), it was not until almost two decades later that the idea was broadened into the modelling of behaviour driven movement patterns across social and geographical spaces (Epstein & Axtell, 1996). Since the 1990s the suite of computational tools available has expanded involving a range of ABM software tools spanning source code languages, model development, and levels of scalability (Abar et al., 2017), making ABM a multi-technology complex product. Such software tools are complemented by ABM libraries and programmes that are used widely in building ABM simulations. While these libraries may not be suitable for modelling complex problems they enable quick execution while also providing a starting point for modelling.

The development of ABM involves cyclical phases of theoretical and empirical analyses. It broadly involves four systematic phases of formulating a real world problem, its computational rendition, executing experimental runs, analysing and documenting the output for iterative development and re-use. Therefore, ABM provides results to a formulated problem using an appropriate simulation development process (Balci, 1994; A. M. Law, 2008; Sargent, 2011). The problem definition identifies the social phenomenon to be simulated as well the objective for such simulation. At this stage the domain to be modelled, the end-user/s who will use the output, and the purpose of the simulation are defined. This is formalised through a conceptual model using theories and assumptions which are then computationally rendered. This involves defining the simulated entity, their attributes, and relationship with each other and the environment while also identifying the programming environment to be used, size and scale of the simulation, nature of input data, distribution, and mobility. This digitisation provides the basis for running the simulation to obtain the results as per the objectives set out in the first instance. The results are then analysed and the process documented for refining the model through evaluation and validation (Heath et al., 2009; Davidsson et al., 2007). Thus, simulation development involves continuous analytical reduction of reality and its technical codification through mathematical abstraction of causal relationships between parameters of interest (Kallinikos, 2009; Chorafas, 1965).

Translating its capabilities in modelling population movements, ABM has been used in contexts of forced displacement. This includes disaster-driven migration incorporating climate change and demographics (Entwisle et al., 2016); influence of climate change on migration in Bangladesh (Hassani-Mahmooei & Parris, 2012) and Burkina Faso (Kniveton et al., 2011; 2012); modelling refugee communities to inform policy decisions (Anderson et al., 2006; 2007); understanding interaction between refugees and military group in refugee camps (Johnson et al., 2009); predicting conflict characteristics and potential conditions and outcomes of the conflict in Syria (Latek et al., 2013); Syrian refugee flows into European countries for policy recommendations on allocation of humanitarian resources (Hattle et al., 2016); capturing social aspects like networks, group formations; and travel distance in forced displacement (Collins & Frydenlund, 2016; Lin et al., 2016). ABM Environment Matrix methodology is one of the ways in which Sokolowski & Banks (2014) propose the development of a ABM for forced displacement. They establish their simulation using an early warning model of forced displacement, match the factors with UNHCR data and develop an assessment template to record model outcomes. The matrix is used for a specific environment such as a Syrian city of Aleppo and run using one hundred replications of the Monte Carlo simulations. This highlights the diversity of resources and knowledge base ranging from modelling, mathematics, and software engineering required to develop an ABM.

Modelling forced displacement requires attention to types of migrants, methods, structuring available data, modelling approach, and the methods of uncertainty associated with data

(Raymer & Smith, 2010). It also becomes important to consider the course of movement of refugees and / or IDPs including when and where they decide to flee and the distance from their first location (Hébert et al., 2018). However, one of the key challenges in modelling forced displacement is data unavailability with regard to features that can explain complex causal relationships within people's movement in the context of forced displacement (Pham & Luengo-Oroz, 2022). Violent conflicts which require urgent modelling for planning and operational purposes present extreme difficulties in terms of access to data. Further, this limited and patchy data is imbued with knowledge shaping the assumptions and selection of parameters which then goes on to structure a slice of its social and operational reality (Jacobsen & Fast, 2019). Thus, model development has to contend with the issue of unstructured parameters and variability in details and assumptions around design, implementation, and documentation. As a result, specifications for such complex products developed for inherently fragile contexts and implemented within globally interlinked humanitarian systems cannot be known in advance. This is because digital products developed for humanitarian management are highly context dependent as needs vary across contexts (Bessant et al., 2014). Context specificity, diversity of functional requirements, and design hierarchies might not be readily observable and available in different forms either as components or interfaces (Murmann & Frenken, 2006). As specialised projects in the absence of mass market conditions, product evolution represents the progress of underlying technologies which requires high levels of architectural knowledge and a system integrator to coordinate among different stakeholders to bring together the required knowledge, components, and resources in the form of specific outcomes (Hobday 1998; Hobday et al., 2003; Davies & Hobday, 2005). Herein, stakeholders not only involve developers and users but providers of components, as well as contextual knowledge and resources (Utterback & Abernathy, 1975; Abernathy & Utterback, 1978).

ABM for forced displacement as a predictive analytic tool can be used to assist governments, multilateral organisations, and NGOs in operational planning by estimating when and where displaced persons are likely to arrive and which camps are likely to become full in the short-term (Pham & Luengo-Oroz, 2022; UNOCHA, 2021). However, the ongoing debate within the ABM community is whether simulations should be used for predictions or for understanding of the phenomenon (Elsenbroich, 2012; Epstein, 2008), particularly given the challenges in obtaining the necessary data for producing a predictive output (Klabunde & Willekens, 2016). The complex and fraught social reality of forced displacement means that there are multiple factors that trigger population movement where a small change of model parameters can result in the significant change in the output highlighting the highly intertwined and integral nature of ABM architecture. This requires tests for sensitivity analysis to validate outputs and execute ensemble runs by simplifying and accelerating key phases of the simulation (Cirillo & Gallegati, 2012).

ABM represents a complex digital product, the development of which depends on the integration of diverse bases of specialised knowledge, resources, and components. As a multi-technology product, it requires high levels of architectural knowledge on part of the system integrator to coordinate amongst different stakeholders. It also comes to be shaped by the extant realities of the context of its application which determines the conditions that structure its development trajectory. From the preceding discussions it can be seen that the development of product architectures, whether modular or integral, are a socio-technical phenomenon. It depends not only on the technical specifications but how those technical specifications can be arranged and organised in a way that enables participation from a diverse range of stakeholder for further product development. While modularity enables such participation through affiliation driven by centralised design rules, integrality involves purposive integration based on specialised knowledge. The following section explores the nature of this socio-technical interdependency as has been studied within IS literature.

2.3 Beyond managerial and economic context: Role of wider conditions of possibility

The role of context in IS research has been much debated with one strand critiquing the over-reliance on context within IS research, often at the expense of theorisation of the technological artifact (Orlikowski & Iacono, 2000). Others highlight the contextual underdevelopment within theoretical models which has tended to exclude any explicit consideration of the context and its many characteristics (Lamb & Kling, 2003; Davison & Martinsons, 2016). This highlights the contention around what counts a relevant context in the study of IS phenomenon (Avgerou, 2019). Digitality and ongoing technological development have led to a shift from formal organisations to non-hierarchical flatter organisational forms like ecosystems. When the formal organisation can no longer be taken for granted within which technological development unfolds, it becomes important to take into consideration the role of the wider environmental context within which IS phenomenon takes place (Winter et al., 2014).

The humanitarian space as a social arena comprises of a multiple actors like donors, UN agencies, governments, technologists, private sector, NGOs, peacekeepers, and military actors (Hilhorst & Jansen, 2010). Technologies structure these relationships and the nature of protection within the humanitarian system (Sandvik et al., 2014). The nature of the underlying technological system shapes the relationship of the actors and activities around it. For example, the transition of UNHCR's refugee registration and identity management system from a closed transaction system to one that facilitated open innovation led to an evolution of activities and incentives among refugees, partner organisations, and third-party service providers (Madon & Schoemaker, 2021). The nature of socio-technical interdependencies around product architectures can lead to the undermining of modes of control embedded within them through novel ways. Iazzolino (2023) demonstrates this in his study of migrant food delivery gig workers who contest their technology driven subjectification and exploitation by registering protest by logging out of the app. Moreover, non-business fields like humanitarian management involve heterogeneous systems of actors, institutions, infrastructure, and data which intersect to bring about coevolution of architecture and ecosystem in the context of its environment (Dawes et al., 2016; Harrison et al., 2012; Bonina et al., 2021). They represent a 'complex, self-adjusting system of resource integrating actors' regulated by enduring norms, rules and values that shape the 'rules of the game' and 'guide how resources are integrated' (Koskela-Huotari et al., 2016, p. 2964). They are structured by activities that involve the acquisition and management of resources, the design and provision of technologies, establishment and implementation of rules of engagement, and response to environmental stimuli (Dedehayir et al., 2018). Relationships come to be shaped by formal and informal rules that structure relationships and the way diverging interests are negotiated, the resources required to bring about a particular innovation, and the nature of relationships between actors (Rush et al., 2021). These are in turn shaped by the relative position of different humanitarian actors, the nature of legitimacy they provide, and requirements from donors. Relationships are structured not just on the basis of resource complementarity but also the ability to manage these relationships and compatibility between different partners in terms of mission, interests, and operational methods (Moshtari, 2016). This becomes particularly important as humanitarian service delivery predicated on digital technologies can often hit bottlenecks due to lack of alignment between service providers and humanitarian organisations along the lines of incentives and activities (Madon & Schoemaker, 2021).

Development of IS takes for granted modes of technical reasoning and acting around digital product architectures, modes of control, and economic advantage (Avegerou & McGrath, 2007). While the digital product innovation literature acknowledges the importance of an environmental fit for successive product development (Tiwana et al., 2010), the role of wider conditions of possibility, beyond managerial and economic ones, enabling such development remains under-explored. Particularly in terms of how they shape the nature of such socio-technical undertaking. Context specific approaches highlight the importance of wider social and historical conditions that shape the processes of technological innovation (Avgerou, 2001; Avgerou, 2000). Context refers to the environing domain that creates conditions of possibility that precipitate in a particular phenomenon (Avgerou, 2019). These are conditions that influence the occurrence of a phenomenon but do not create it (Elwick, 2012; Hacking, 2002). Therefore, the occurrence of a phenomenon depends on causal trajectory or the movement of causal effects on an entity and causal autonomy or the causal effects between individuals and technology (Markus & Rowe, 2018) where the relationship between them depends on the contextual environment. Both act on contextual conditions (factors within the environment of the IS phenomenon shaping its outcomes) and through relations that associate an IS phenomenon with the conditions of its environment (Avgerou, 2019).

The design and development of technologies for humanitarian management need to pay attention not only to the product but also the complex humanitarian system through which such products are to be deployed (Nielson et al., 2016). Developing humanitarian technologies require attention to multiple socio-technical components like the computational architecture in the form of hardware and code; the context of crisis, ethics, and principles; stakeholders like donors, developers, and humanitarian organisations; and knowledge and skills that determine capacities for development, deployment, and implementation (Akter et al., 2021).

The contextual environment can operate both as a material resource environment as well as a set of formal and informal rules (Constaninides & Barrett, 2006) which determine the selection of resources and how they are combined and deployed (Greenwood et al., 2002). The environment can have an endogenous influence in the form of actions such rule systems legitimate, systems of governance they establish, and resources they make available that shape the internal design choices (Scott et al., 2000; Tiwana et al., 2010). Exogenous environment pressures can be in the form of the social and regulatory pressures which seek to influence existing systems (Zucker, 1983). This necessitates a multi-level analysis to understand how societal, inter-organisational, and individual factors are interrelated with the material resource environment and conditions of possibility therein (Currie & Guah, 2007). This becomes particularly important as existing literature on digital product innovation remain limited to interdependent dynamics between the technical architecture and its proximate context of the ecosystem. Wider conditions of possibility i.e. the distant environment in terms of underlying technology developments, improved resource capacity, social conditions, and policy and regulatory contexts that structure the innovation trajectories have often been neglected within its study (Alaimo et al., 2020; Avegrou & McGrath, 2007; Zittrain, 2008; Jacobides et al., 2018; Bonina et al., 2021)

As a result, development of technologies follow a trajectory that is a by-product of negotiations between different stakeholders which are in-turn shaped by partial attainment of goals or by privileging the attainment of some goals over others (Heracleous & Barrett, 2001; Sabherwal & Newman, 2003) subtended as they are within complex technological systems and the modes of action they enable and constrain (Kallinikos, 2009; Alaimo et al., 2020). Contextual research helps to arrive at system level outcomes by identifying influences from conditions at the level of manifestation (Avgerou, 2019). As a result it helps in the study of broader and more complex socio-technical arrangements implicated in the development of complex products like ABM.

2.4 Problematisation and research questions

The extant literature on digital product innovation highlights how the generative nature of digital technologies pervades through product architectures and how product architectures organise their innovation trajectories. However, theory development has remained predicated on a particular type of architecture i.e. modular architecture and the way it shapes and coevolves with its proximate context of the ecosystem. Layered modular architectures engender ecosystem by affiliation based on standardised interfaces and serendipitous development driven by the digitality and generative capabilities. It does not adequately explain the process of developing complex digital products like ABM which present highly intertwined architectures that require specialised integration of knowledge, resources, and components making coevolution processes more purposive. Complex product architectures are developed under different conditions and the mode of their innovation exhibits high reliability on the external environment to which it continually attempts to respond. While some studies have acknowledged the implications of wider environmental conditions on the product development process, such distant contextual conditions have not received adequate attention within theorisation on digital product innovation. This becomes particularly important in case of system integration activities to coordinate the diverse and specialised requirements for complex products. This research aims to develop a contextualised theory of digital product innovation by exploring the process of developing complex digital products and how it comes to depend on the trilateral interdependency between the architecture, ecosystem, and environment. Towards this end, the study engages with the following principal research question:

How do technical and contextual interdependencies shape digital product innovation of complex products like the Flee ABM?

The aim of the primary research question is to develop a holistic theory of digital product innovation by explicating the relationships between the generativity of digital product architectures and conditions within its proximate and distant context.

This main question is answered with the help of two operative research questions:

(1) **How do complex digital architectures and their development ecosystems coevolve?**

The first step towards developing a holistic theory of digital product innovation involves explaining the relationship between the complex digital product architectures and their development ecosystems. These explanations proceed from the description of the nature of architecture and ecosystem observed and how their particular characteristics underpin ongoing product development.

And

(2) **How are environmental conditions interrelated with the coevolution of the complex digital architecture and the ecosystem?**

The second step involves incorporating the explanations arrived at through the previous research question and exploring them within the wider environmental conditions identified to understand how product development proceeds through this trilateral interdependency between the architecture, ecosystem, and environment.

3. THEORETICAL FRAMEWORK

Chapter summary: *This chapter develops the adapted choice framework as an analytical approach to unpack socio-technical interdependencies within complex digital product innovation through contextual considerations. The adaptation of Kleine's Choice Framework involves a synthesis of theoretical imperatives from the literature and contextual concerns of the socio-technical approach. Underlying assumptions about causal structure within such an approach is explicated and domain of enquiry and relevant contexts are delineated. Once the contextual elements have been defined, the adapted choice framework helps develop explanations for the causal trajectory of digital product innovation. The framework helps draw attention to the opportunity structure of the resource portfolio, nature of relationships in the proximate context of the ecosystem, and the wider technological, social, economic, and political conditions in the distant context of the environment. These in conjunction with the digital product architecture help understand their implications for the overall process of complex digital product innovation.*

3.1 Socio-technical interdependencies in digital product innovation

The literature review helped highlight the strands of literature on which theorisation for digital product innovation is built. While the product innovation literature helps understand how innovation is organised (Baldwin & Clark, 2000; Clark, 1985; Utterback & Abernathy, 1975), generativity as an inherent quality of digital products facilitates distribution of innovation among a wider ecosystem of participants than was previously possible (Yoo et al., 2010; Zittrain, 2006; 2008; Benkler, 2006). Therefore, digital product architectures and ecosystem can be derived as the key constructs or building blocks for theory development in this domain where the architecture organises innovation activity and ecosystems negotiate opposing tensions that mutually shape the innovation process (Eaton et al., 2015; Ghazawneh & Henfridsson, 2013; Tilson et al., 2010; Wareham et al., 2014). However, extant literature is premised on a particular type of architecture i.e. modular architecture and how generativity augments or extends it (Yoo et al., 2010). Underlying this theory development is the assumption of a mature product where product features are known in advance which can then be decomposed through standardisation resulting in product variation and production at scale (Clark, 1985; Salvador, 2007). However, as discussed, development of complex products exhibit alternative organisation for innovation since they tend to operate with tightly coupled integral architectures and high degrees of uncertainty which makes their innovation dynamics differ from mass produced ones (Ulrich, 1995; Hobday, 1998; Hobday et al., 2000). Due to successive multi-actor configurations, complex product development needs to be responsive to changing requirements from users who come from different contexts, changes in underlying technologies, policy and regulatory pressures among others. Existing literature acknowledges some aspects of such wider environmental concerns. Jacobides et al. (2018) and Bonina et al. (2021) highlight the importance of regulators and pressure groups, Zittrain (2006) demonstrated how external conditions like cyberattacks present significant innovation dilemmas, Alaimo et al. (2020) highlighted how broader technological developments propelled the innovation dynamics of TripAdvisor in conjunction with its evolving internal capacities. However, these broader environmental concerns have not received adequate attention for theorisation. While contextualist research in IS highlights the importance of multi-contextual analysis for IS innovation (Avgerou, 2000; Avgerou, 2001; Avgerou, 2019, Avgerou & McGrath, 2007), theorisation within digital product innovation has remained limited to studies of the technical architecture and its proximate context of the ecosystem. Given the nature of digital product development and the role of wider environmental conditions, it becomes important to understand the interdependencies not just between the digital architecture and its proximate context of the ecosystem but look at these dynamics in conjunction with the distant context of wider environmental conditions. This would help provide a better understanding of the conditions underlying architectural design choices that propel the innovation process through each successive stage of development.

Theorisation on IS innovation has tended to be perspective-centred where concepts and assumptions from other research streams have been used as a foundation for theory building which precludes a more holistic view of the phenomenon. A more holistic process oriented approach to theorising about IS innovation helps provide a structuring device for understanding the phenomenon (Fichman et al., 2014). Acknowledging multiple planes of influence or context implicated within a given phenomenon helps develop holistic explanations for the same (Pettigrew, 1997). Enduring indeterminacy of digital products conferred by their generative nature means newer products continue to acquire layers of inheritances which shape generative systems (Kallinikos et al., 2013; Kallinikos, 2009; Faulkner & Runde, 2019). Generative systems open up participation for a diverse range of actors to converge in the innovation process inviting multiple contextual influences. Therefore, digital product innovation comes to be mutually shaped by the specific character of the technology and its architecture, the proximate or inner context of its development ecosystem with its diverse range of participants, and the outer or distant context of the environment involving economic, political, social, and sectorial conditions (Pettigrew, 1997). With digital technologies changing modes of organising, this requires the need to develop appropriate approaches for understanding the relationship between newer digital phenomenon and their relevant contexts (Avgerou, 2019). This would help in understanding how particular environmental and actor configurations coincide with digital product architectures and systems (Alaimo et al., 2020) by explicating the conditions that underpin design choices at each successive stage of innovation. Endogenous design choices represent actions embedded in contexts which shape their information, insight, and influence (Pettigrew, 1997). Unpacking these diverse and heterogeneous conditions across layers of context in relation to a digital phenomenon requires the synthesis of IS contextualist approaches into a framework that can provide a flexible structure for multi-level analysis. Kleine's Choice Framework is a holistic and systematic approach to understand how given resource portfolios in relation to interpersonal relationships and wider environmental conditions result in opportunities that enable the materialisation of a given choice (Kleine, 2010; 2011; Sen, 1999). While contextualising approaches help in explicating the conditions around the phenomenon, Choice Framework helps bring them together in developing holistic explanations through multi-level analysis that link action and conditions to outcomes. Since the Kleine's framework was developed for a different context of application (ICT4D), this study adapts the approach by synthesising theoretical imperatives for studying socio-technical interdependencies and digital technologies.

3.2 Considerations for contextualising digital product innovation

Within this research, context is taken to mean the environing domain around an IS phenomenon, the exploration of which helps in explaining such phenomenon (Scharfstein, 1989; Avgerou, 2019). The understanding of context as an environing domain contains important assumptions about the nature of causality and causal explanations. Markus & Rowe's (2018) framework of causal structure put forward three dimensions of causal assumptions: causal ontology, causal trajectory, and causal autonomy. Causal structure refers to the researcher's assumptions about the nature of causal influences in IS phenomenon which becomes important to explicate within a theory building endeavour given the diversity of theoretical approaches within the domain. However, the conception of causality itself has contentious philosophical strands as to what constitutes causality i.e. whether it is variables, actors, or events, whether it can or cannot have multiple causes, feedback loops, bidirectional or multidirectional effects (Markus & Rowe, 2018). Drawing on Lakoff & Johnson's (1999) theory of embodied realism, Markus & Rowe (2018) put forward a pluralistic approach to thinking about causality in their framework of causal structure. A pluralistic approach avoids the pitfalls of extreme idealism or relativism. It is based on the premise that reasoning about the world is based on experience and interpretation as a result of which concepts of causation become important ways of planning and acting in the world. Concepts of causation can be described as thin or thick where the former provides the essential criteria in the form of data to develop causal claims and the latter builds on the former to provide higher level descriptions that enable analysis and inference (Cartwright, 2004). Developing thick causal concepts helps in moving from data to descriptive empirical statements (Lee & Baskerville, 2003). Within this study, developing thick causal concepts provides one of the building blocks for theorising the phenomenon of digital product innovation as transmutation of different conditions that result in successive design choices as a function of the interdependency between the digital architecture, ecosystem, and environment.

Causal ontology refers to the researcher's view on whether causality is real, drawing on Avgerou (2019) this research takes the view that conditions environing the IS phenomenon creates the possibility for their occurrence but do not directly cause it i.e. the occurrence of the phenomenon depends on but is not determined by the conditions in its context (Elwick, 2012; Hacking, 2002). Causal autonomy refers to movement of causal effects between social actors or elements and technology (Markus & Rowe, 2018). The definition of context as an environing domain does not involve *a priori* assumptions about causal autonomy but depends on the approach taken to theorise this relationship. This study takes a socio-technical perspective of causal autonomy to understand a given phenomenon as mutually shaped by social and technical elements. Causal trajectory involves the movement of causal effects on an affected entity where the entity is foregrounded as one that is changing or 'moving in space and time' (Markus & Rowe, 2018, p. 1259) where the effects of the movement are relegated to the background as conditions that explain the changes that the entity is undergoing. The causal trajectory of complex digital product innovation is explored in this research as a function of its digital architecture, ecosystem, and environment. However, developing contextual theory involves layers of explication beyond the causal structure.

On looking for causality in contextual research Pettigrew (1997) suggests that contexts provide the scope of examining causal processes directly. Developing causal explanations help understand why and how a phenomenon occurs (Avgerou, 2013). Tightly interconnected systems can lead to non-linear and highly unpredictable forms of causality (Grabowski & Roberts, 1999). Arguments against incorporating causal explanations stem from the concern that they cannot fully capture the nature of IS phenomenon and how they are brought about. Particularly because they cannot fully account for the interpretive flexibility of actors encountering technology since the processes shaping IS phenomenon are dynamic and largely unpredictable (Walsham, 1995; 2006; Markus & Robey, 1988). However, drawing on Giddens (1984) Avgerou (2013) highlights the importance of causal explanations because they underpin generalisable theoretical concepts based on analysis with causal explanations developed through a combination of general theory and analysis of observed phenomenon (Salmon, 1998). Causal explanations help form the link between initial state and observed outcomes within a process by identifying mediating conditions (Avgerou, 2013). Just as thick causal concepts help in developing descriptive empirical statements from data, mediating conditions become the basis for translating descriptive statements to theoretical statements through analysis (Lee & Baskerville, 2003; Avgerou, 2013). As a result of which causal explanations have assumed increasing importance within contextualist theory and theorising (Avgerou, 2019) which aims to explain the transmutation of diverse multilevel contextual conditions into action and outcomes (Pettigrew, 1997). Contextualist theorising does not involve looking for individual causes but diverse conditions of possibility across multiple levels or ‘constellation of forces’ that shape a particular process and outcome. As a result, the search is for proximate not final causes or the multiple intersecting conditions that culminate in particular outcomes (Tilly, 1984; Ragin, 1987). This involves an assumption of circular reciprocating modes of causality as opposed to uni-directional cause-effect relationships that can be “tested via hypothetic-deductive logic” (Orlikowski & Baroudi, 1991, p 9). Consequently, causal explanations derived within contextualist research are necessarily partial and indeterminate i.e. without predictive value but help provide theoretical insight which enable transferability of such insights for future studies (Lee & Baskerville, 2003; Eisenhardt, 1989).

While the theoretical approach helps explain assumptions about causal structure, contextualist research also involves delineating the domain of enquiry and what counts as the relevant context which serve as important aspects of theory development. Once these elements have been outlined, there needs to be analytical approach that helps explain the causal trajectory under investigation. The following sections unpack these elements in relation to this research while positing the adapted choice framework as the approach which, in contextualising the relationships and interdependencies under consideration, helps explain their role in the phenomenon being studied.

3.2.1 Theories of technology

Theories of technology explore the relationship between technology and change. IS literature spans a diversity of theoretical approaches which can be summarised as the socio-technical perspective, the actor network theory, interaction perspective of socio-materiality, and intra-action perspective of sociomateriality (Avgerou, 2019). According to the socio-technical perspective, IS phenomenon is simultaneously technical and social (Sawyer et al., 2003). Analysis from a socio-technical perspective must address both these aspects as well as the reciprocal relationships between them. A socio-technical perspective attempts to overcome the limitations of viewing IS phenomenon either as technologically or socially deterministic (Doherty & King, 2005; Robey et al., 2001; McLeod & Doolin, 2012).

Inter-action perspective of socio-materiality looks at localised experiences of human and material agency shaped by the human capacity to act and capabilities offered or afforded by technologies (Faraj & Azad, 2012; Markus & Silver, 2008; Zammuto et al., 2007; Leonardi, 2011). In contrast to the inter-action perspective which holds the IS artefact and human actor to be ontologically separate with each possessing causal capacity, the intra-action perspective of sociomateriality, following Barad's agential realism, sees IS phenomenon constituted by inseparable physical and material entities where such phenomenon emerge as a part of their 'entangled intrarelating'(Barad, 2007, p. ix; Orlokowski & Scott, 2008). Actor network theory arguably rejects the notion of context where IS phenomenon can be construed as networks of actors that can include both technological artefacts as well as individuals and collectives where the performance of relationship between these heterogenous entities continuously assemble such IS phenomenon (Latour, 2005; J. Law, 1991). Underlying theories of technology are foundational theories of action which provide insights on the IS phenomenon and broader domains of context (Avgerou, 2019). Theories of action contain assumptions about how individuals act in response to their environment and how such action can be explained through a given theory's ontological position. Theories like actor network theory arguably reject the notion of social structures in favour of a relationship-based approach while the intra-action perspective of sociomateriality views agency as enacted through sociomaterial structures which are constantly reconfiguring each other. IS literature has extensively drawn on structurational perspectives of action where participants within an IS phenomenon are knowledgeable actors who reflexively act in relation to other social and material aspects in their environment where their actions are both reflective and practical (Giddens, 1984, Avgerou, 2019; McLeod & Doolin, 2012; Engestrom, 2004). This study too is premised on this approach in taking a socio-technical perspective of causal autonomy where IS phenomenon comes to be mutually shaped by social and technical elements. An acknowledgment of socio-technical interdependency is keeping in line with the pluralistic approach to causality in avoiding social or technological determinism. While this helps to incorporate wider conditions of possibility beyond techno-managerial ones, given the multiple planes of influence and generative nature of digital technologies implicated in the process, Kallinikos et al. (2013) question the more relativistic approaches to theorising about technology. This is because, they argue, technological developments are product of 'wider and time-ridden' conditions that go beyond the local context and localised enactment and interpretation (p. 367).

The choice of theory has important implications with relation to assumptions about ontology, scope, and scale (Avgerou, 2019). Theoretical approaches where context is analytically separate from the IS phenomenon under study i.e. the socio-technical perspective and inter-action perspective of socio-materiality corresponds to distal perspective that assumes a world made of stable material and social entities. However, despite sharing a similar approach to context, the theories can differ in the scope of analysis. For example, while the socio-technical perspective enables the exploration of context beyond its immediate setting, the inter-action perspective of socio-materiality does not. Actor network theory and the sociomaterial approach correspond to the proximal perspective where the world is in a dynamic state of making as a result of interacting sociomaterial entities which continue to reconfigure the context. The use of the term proximate context in this thesis is distinct from the proximal ontology described herein where proximate context refers to the conditions in the immediate setting of the phenomenon under study.

The selection of the theoretical approach also determines the scope and scale of contextual IS enquiry in terms of the variety of conditions of possibility and magnitude of the domain of enquiry (Avgerou, 2019). This foregrounds the tension between scale and detail within contextual IS research. Research that considers context as the immediate domain is often critiqued for not being adequately contextualised (Kallinikos, 2004; Pollock & Williams, 2009) while those that consider larger domains and longer histories present an oversimplification of the causal relationships (Knorr-Cetina, 1981). Layered contextual strategy enables to go beyond the immediate setting to explore larger conditions of possibility (Madon, 1992; Walsham, 1993; Pettigrew, 1985) while relational contextual strategy entails detailed analysis of the network of connections among internal participants of an IS phenomenon and their relation to other individuals, collectives or artefacts (Kling, 1987; Kling & Scachhi, 1982).

Contextual research is always partial in scope i.e. each study only investigates only some of the multiple conditions in the surrounding context of the phenomenon (Townley, 2008). With IS theorisation being rarely purely inductive, conditions factored into theorisation are informed by middle-range theories (Avgerou, 2019). The literature elicits some aspects of the relationship between the key constructs of architecture, ecosystem, and environment which are explored in this study. However, this research aims to explicate the wider diversity of conditions and their interdependent relationships through the contextualisation approach of the adapted choice framework. Moving towards this approach involves understanding and delineating the domains of enquiry and the relevant context.

3.2.2 Domain of enquiry

The predominant view of context is that of a social domain which contrasts with existing approaches to theorising technology within IS (Avgerou, 2019). Given, the socio-technical approach of this thesis and operating within insights derived from the literature this requires the explication of social and technical element for analysis. As mentioned earlier, the literature highlights the role of the architecture in organising innovation. While the layered modular architecture distributes design tasks among a diverse and heterogeneous group aided by generative capacity of digital technologies, complex product development proceeds through carefully crafted coalitions of diverse range of stakeholders (Yoo et al., 2010; Hobday, 1998; Hobday et al., 2000). Digitality and scope of digital product development requires us to look beyond traditionally studied domains like organisations, industries, countries, regions, and communities since organisational forms engendered by digital product architectures cannot be slotted into *a priori* assumed social collectives (Avgerou, 2019). As digital product development increasingly happens in flatter non-hierarchical organisational forms like ecosystems (Yoo et al., 2010; Jacobides et al., 2018; Ghazawaneh & Henfridsson, 2013; Eaton et al., 2015), it becomes important to understand the nature of ecosystem and the structure of the roles and relationships in relation to the digital product architecture.

As complex digital products like ABM are highly specialised and are developed in non mass-market conditions they involve a diverse range of participants beyond the traditional user, developer, and manager groups (McLeod & Doolin, 2012). They include a range of stakeholders with specialised knowledge, ownership, and control over the different resources required within the product development process (Hobday, 1998; Hobday et al., 2000). As a result, their actions come to be structured by the opportunities and constraints offered by conditions both within its inner and outer contexts as well the technical architecture (Pettigrew, 1997). These function as conditions of possibility to understand, rationalise, and legitimate particular courses of action. However, the structuring conditions of the environment do not operate in a deterministic way but form an order of rules and resources that shape the feasibility and appropriateness of actions within a given context (Jones, 1999). However, the scope of actions are circumscribed by the material capabilities and the actual demand of the technology in question. Outcomes of actions have the potential to affect future action by shaping structure, actors, task, and technology (McLeod & Doolin, 2012; Leavitt, 1964). Conceptualising the proximate context of ecosystems as a structure of interdependent relationships helps understand the nature of socio-technical interdependencies therein as opposed to the actor centric view predominant in studies of layered modular architecture of digital technologies that focus on affiliation between stakeholders (Adner, 2017). This structure of relationships requires configurations of action, actors, position, and links that shapes the nature of interdependence that can arise in cross-contextual settings of proximate and distant contexts (Alaimo et al., 2020; Adner, 2017).

The ecosystem as affiliation approach sees ecosystems as association of actors defined by their affiliation to each other and the focal actor or product owners within layered modular architectures (Adner, 2017; Alaimo et al., 2020). It is concerned with the number of partners, network density, and centrality of actors in larger networks. It focuses on increasing the number of actors that can connect to the focal actor thereby increasing its centrality and power while also increasing opportunities for new interactions and serendipitous development. However, the concept of ecosystem as affiliation tends to look at aggregates without unpacking the specifics that result in the transmutation of ecosystem activities into ongoing product development. Ecosystem as structure begins its observation from the focal innovation, in this case, the product to be developed which is more in line with the innovation trajectories of complex products (Hobday, 1998; Hobday et al., 2000). Members have defined positions and flows of action and product development comes to depend on the extent to which socio-technical configurations are able to resolve attendant design problems. This comes to depend not just on economic incentive structures and individual stakeholder interests but also the way existing technical and contextual conditions reconfigure how such innovation comes to be constituted.

3.2.3 Relevant context

Contextualising domains of enquiry involves layered and relational approaches (Avgerou, 2019). According to the layered view, social collectives emerge from the constitution of their subsystems. It uses cross level analysis to find congruency between the factors and processes within a social collective that shape an IS phenomenon. This cross-level analysis often involves an out-contextualisation whereby the researchers look for conditions of possibility within broader and more complex systems than the one under observation (Madon, 1992, Walsham, 1993; Pettigrew, 1985). In contrast, the relational approach to contextualisation explores the relationship between internal participants of an IS phenomenon and other stakeholders or even actors that exert influence on their actions. This does not assume differentiated cross-system level analysis and *a priori* discrete entities like the formal organisation. Instead, it looks at interdependent stakeholder groups involved in IS development and use as well as social actors beyond the immediate IS phenomenon like professional associations, funders, auditors, and regulators etc. (Kling, 1987; Kling & Scachhi, 1982). The two approaches to contextualisation highlight the need to delineate the extent to which conditions beyond the immediate setting of the IS phenomenon is to be admitted within the ambit of analysis (Avgerou, 2019).

System integration of complex digital products is about computational design, organisational arrangements, and management and negotiation of uncertain conditions that structure their ongoing development. Much depends on the relative power exerted by component providers and the importance of those components within the overall system. Choices around system design then come to depend on the overall relationships with the component producers. Because complex products are not mass produced for final consumers but tailored for institutional consumers, these intermediate customers are intimately involved in the design and development process (Hobday, 1998). The innovation idea often originates from these consumers in the form of a pre-defined need where ancillary services like finance and training become a part of the complex product ecosystem. This requires a rethinking of management processes and best practices that have originated from consumer goods produced in volume. The nature of the specific product or the system is an important determinant of the type of organisational form supporting its iterative development (Yoo et al., 2010). The nature of component inputs, complexity of component interfaces, range of knowledge and skills involved, intensity of user involvement together determine the overall nature of complex digital product innovation. The nature of the artefact and the process of its design and development become inextricably intertwined and it becomes difficult to determine one without the other. Further, the external environment of rules, norms, and regulations are implicated within the coevolutionary trajectory of digital product architectures and ecosystems. This highlights a need to think about multilateral inheritances along with multilevel planes of influence that do not fit neatly into pre-existing contextualisation strategies like layered or relational as they incorporate characteristics of both. While this study's delineation of the relevant context as proximate and distant helps factor in the multi-level influence across different layers of context, the proximate context consists of distributed arrangement of diverse and heterogeneous actors linked together in the development of the product. This does not reflect an ontological elision but acknowledgment of the spirit plurality of causal structure and circular, reciprocal effects of multiple contextual influences (Pettigrew, 1997).

3.2.4 Theorising the relationship between context and phenomenon

The diverse nature of socio-technical interdependencies means such relationships between social and technical elements might not be fully reconciled and can involve inter-locking constraints and loose coupling (Winter et al., 2014). The phenomenon under consideration in this study i.e. digital product innovation is embodied in the digital product, the architecture of which organises the proximate or inner context of the ecosystem. The ecosystem negotiates contending tensions to advance the digital product development thereby propelling the innovation process. With ecosystems composed of heterogeneous stakeholders, they derive purpose, meaning and structure from the multiple contexts in which they are embedded and which in turn determines their resource ownership, control, and contribution. This results in multilateral, recursive inheritances from multiple contexts requiring attention towards wider conditions of possibility in the distant or outer context (Winter et al., 2014; Pettigrew, 1997). As stakeholders like the system integrator, infrastructure provider, humanitarian organisations etc. inherit purposes and meanings from multiple contexts, this leads to upward causation where the whole becomes more than the sum of its parts (Winter et al., 2014). Multilateral inheritances with differential ownership and control of critical resources results in constant negotiation and compromise of system level goals. This constant negotiation of goals and actions become important to understand at various stages as influences from multiple contexts are managed within a given design choice. Due to the reciprocal and iterative nature of this relationship, it becomes important to explicate how these diverse conditions intersect to give rise to a particular choice in the form of functional extensions of products, and particularly how these conditions change and co-evolve over time.

Privileging context within a socio-technical perspective involves acknowledging mutually shaping technical and social systems. The multi-disciplinarity and specialisation required to build complex digital products goes beyond the organisational container to think about stakeholder dynamics within an ecosystem containing multi-directional planes of influence (Winter et al., 2014). This decentres the traditional approach to IS where organisations were supposed to provide the overall goals and create technical systems that provided the context for technology development and use. Generativity and digital technology have opened up the space for distributed development and enabled the creation of large cyberinfrastructures that are accessed by diverse communities. Moving beyond ‘organisations as a container’ to think about distributed development highlights the need to explicitly consider the particular conditions of possibility that give rise to a given phenomenon. It allows the consideration of multiple conditions of possibility within the analytical ambit to acknowledge and approximate causal trajectories that shape digital product development. This mode of organising for innovation does not embed unilateral managerial imperatives but have to take into account the balance between opposing tensions like generativity and control, stability and evolvability, and other tensions that arise during the process of innovation (Eaton et al., 2015; Wareham et al., 2014; Tilson et al., 2010). Keeping in line with the socio-technical approach, the dynamic and mutual interplay between the social and the technical jointly generates outcomes with implications for mutual causality that involves joint optimisation of social and technical elements with four interacting variable classes - two in social system (structure and actors) and two in the technical system (technology and tasks) (McLeod & Doolin, 2012; Leavitt, 1964). Contemporary IS development includes software and hardware vendors, outsourced contractors, external project managers, and consultants. They act purposefully within defined roles and relationships with differential access to material and non-material resources in terms of knowledge, skills, expectations, interests, and values shaping their actions. Their actions are further informed by their interpretation of the task at hand in terms of project scope, scale, complexity, and available resources. The task is further determined through the architectures as well as project deliverables such as specifications for proposed IS solutions. Actions are also structured by the opportunities and constraints offered by structures and properties of the context i.e. the immediate organisational context and the wider socio-economic context within which the organisation is located. These function as sources of conditions of possibility for actors to understand, rationalise, and legitimate particular course of action. In this way they are shaped by formal and informal organisational structures of relations, authority, norms, and rules (Orlikowski, 1992; Knights & Murray, 1994). However, as highlighted earlier, the structuring conditions of the environment do not operate in a deterministic way but form an order of rules and resources that shape the feasibility and appropriateness of actions within a given context. However, the scope of actions are circumscribed by the material capabilities and the actual demand of the technology in question. Outcomes of actions have the potential to shape future situated action in material and non-material domains by shaping the environment, actors, task, and technology.

Taking into account multiple interconnected levels of context and reciprocal relationship between context and digital architecture enables a multifaceted explanation of change as digital product development happens over time rather than as a linear or singular phenomenon (Pettigrew, 1987; 1990; 1997; Pettigrew et al., 2001). The digital product development process provides the opportunity and site for action and interaction between internal and external contexts in relation to the technical architecture. Acknowledging the interdependency between digital product architectures, ecosystems, and environment overcomes the inherent limitation in viewing digital product innovation simply as a technical-rational process or a purely social process that privileges the interactions between actors to the exclusion of the technology at stake or vice versa. By bringing both the social and technical within analytical purview, it helps to understand the reciprocal relationship between them when clear boundaries between the social and the technical are less clear cut (Orlikowski & Iacono, 2001). It helps recognise that digital product development outcomes emerge from iterative reciprocal relationships between digital architectures, ecosystem, and the environment which shape the process.

3.2.5 Rationale for choice of theory

As mentioned earlier in the chapter, concepts and assumptions from different research streams have been used as a foundation for theory-building in IS innovation (Fichman et al., 2014). Different disciplinary areas have informed the study of digital product innovation drawing from information systems, strategy, management, organisation studies, and economics that have led to a diversity of underlying conceptual approaches for the study of the phenomenon. This has led to an absence of a shared vocabulary and a coherent set of theoretical frameworks which hinders the development of theoretical and conceptual tools for understanding the phenomenon (Nambisan, 2018). However, broadly two perspectives can be discerned within the literature i.e. the architecture and the ecosystem perspective (Nambisan, 2018). The architectural perspective takes the product as the unit of analysis to understand how different digital components or artifacts and their interconnections determines the product's innovation trajectory. While architecture is understood to organise the innovation process among a heterogeneous set of actors (Yoo et al., 2010), it precludes a more dynamic view of actors and agency such as their roles and relationships and implications for digital product innovation. Within the ecosystem perspective, the architecture is largely relegated as the background condition for ecosystem emergence while the management of opposing tensions or paradoxes among product owners takes precedence (Wareham et al., 2014; Eaton et al., 2015; Tilson et al., 2010). Even in their attempt to simultaneously take into account the 'product view' and the 'social interaction view' of ecosystems, Fürsentau et al. (2023) look at the isolated impact of user base growth on one or the other rather than relating each perspectives' underlying assumptions to each other, as discussed earlier. Different perspectives have differing emphasis on the focal unit of analysis. As a result, taking architecture and ecosystem jointly into account within the phenomenon of digital product innovation results in use of conceptual vocabulary from the different theoretical traditions. The way these have been used in the literature helps explain an aspect of the phenomenon under consideration i.e. either architectural or ecosystem priorities rather than a synthesis of both. However, the socio-technical approach highlights the joint optimisation of both the technical and social indicating the simultaneous emphasis on both the architecture and the ecosystem and the relationship between them to explain the process of complex digital product innovation. This also helps overcome the under-socialisation and over-socialisation dichotomy within contextual research by placing simultaneous importance on both (Avgerou, 2019). In parallel, taking into account the role of wider environmental conditions, as suggested by contextualist research and complex product innovation, entails accounting for multilevel influence across layers of contextual conditions such as the inner or proximate context and outer or distant context of the ecosystem and environment respectively (Lyytinen & Newman, 2008; Pettigrew, 1997). This highlights the rendering of underlying imperatives from different bodies of literature in developing a theoretical framework for this study.

Preceding discussions in earlier sections highlighted how contextual research involves a negotiation between scale and detail with the need to incorporate both layered and relational strategies within contextualising approaches for digital product innovation. This is to simultaneously incorporate both multi-level influences across layers of context as well as the distributed arrangement of diverse and heterogeneous actors linked together within the innovation process through the digital architecture as the extant literature indicates. Moreover, innovation in complex digital products have to contend with the requirement of overall specialised knowledge of the product and the distributed open-endedness engendered by digital technologies (Lyyra, 2018). However attempts to theorise the relationship across different layers i.e. between technology and action of different actors within the proximate context as well as technology, action, and the wider contextual conditions beyond the immediate domain need to negotiate the enduring tension of incorporating wider set of conditions without compromising the level of detail.

This raises questions about what could be an appropriate explanatory device. A layered approach is akin to a system theoretic stance which takes into account national, international, or even industry or organisation context beyond the localised enactment of the phenomenon (Avgerou, 2019; Brynjolfsson & Hitt 2000; Madon, 1992, Walsham, 1993). However, such an approach would not help adequately understand the interdependent nature of relationships within the ecosystem which presents flatter non-hierarchical organisational forms enabled by the generative nature of new digital architectures (Avegerou, 2019; Yoo et al., 2010). This is because tracing influences from larger domains beyond the immediate setting of the phenomenon present much less detail about the same and interdependencies therein (Knorr-

Cetina, 1981). On the other hand, limiting the process of innovation to the sum of interdependent relationships ignores the role of technologies in organising actors within the innovation processes and role of wider contextual conditions. Conditions enabled by the nature of generative digital technologies facilitate distributed, product-agnostic and proliferating development (Yoo et al., 2010). Technological elements like standardised digital interfaces become the mode of managing ensuing tensions among actors in the innovation process within the resultant ecosystem (Ghazawneh & Henfridsson, 2013).

One of the theoretical approaches to unpack relationships at the human-technology interface has been the theory of affordances which helps analyse the relationship between technology and action in the context of their immediate setting (Cecez-Kecmanovic et al. 2014; Faraj & Azad, 2012; Leonardi, 2012). Technological affordances become the possibilities for goal oriented action based on the cognitive interpretation of an actor or group of actors (Faik et al., 2020; Faraj & Azad, 2012; Fayard & Weeks, 2014; Markus & Silver, 2008). However, more recent work on affordances has tried to go beyond the immediate setting of a phenomenon and attempted to link material affordances of technologies to wider social change. This involves translation of individual cognitive perception of technologies into collective ones leading to macro-level conditions driving social change (Faik et al., 2020). Faik et al. (2020) propose future research linking localised approaches of technology-in-use to wider social conditions by using socio-cognitive aspects of the relationship between an individual and the technology as explanatory constructs. Faik et al. (2020) aim to explain the link between affordances of technologies-in-use and social change based on technical artefacts imbibing social norms over time. However, the role of how wider social conditions, particularly when they involve a set of heterogeneous actors from different contexts, impinge on the development process still remains to be explored (Winter et al., 2014; Kallinikos et al., 2013). Moreover, using a cognitive theory like affordances to understand wider conditions of change would mean identifying long and complex causal chains from the level of individual cognition upwards which runs into methodological challenges in explaining wider socio-technical phenomenon (Markus & Robey, 1988; Avgerou, 2013). Further, it aims to extrapolate from the role of ‘technology in use' and their transmutation into ‘collective affordances’ wherein technologies assume the nature of infrastructures which then become the foundational basis for social processes (Faik et al., 2020). IT affordances refer to users’ perception of possibilities of action in relation to the IT artefact as individuals reconcile their own goals with respect to it (Markus & Silver, 2008; Leonardi, 2012). Given the role of affordances within IT-in-use, it could be argued that it comes into play when users encounter relatively mature technologies with stabilised functionalities. However, the development of digital products, particularly complex ones, involves the dynamic conceptualisation of evolving product functionalities. Moreover, developing technologies go beyond reconciling action with respect to technological possibilities, it also involves anticipating and understanding user needs and incorporating them into the design process. Clark (1985) highlights how in the initial phases developers need to select from available technological choices in the absence of adequate understanding of user needs or preferences within the wider social and economic environment. Hanseth & Lyytinen (2010) echo similar concerns wherein designers need to work towards completeness of their design within evolving user demands. Further, in conditions of distributed development enabled by digital technologies, even more so in the development of complex products, developers need to simultaneously manage user needs as well as other stakeholders on whom the overall development of the product depends.

Yoo et al. (2010) proposed the new organising logic of digital product innovation involving heterogeneous actors pursuing their own innovation strategies which reciprocally and recursively influence each other through wakes of influence (Boland et al., 2007; Eaton et al., 2015). This is because such new organising logic is premised on innovation being doubly distributed on account of digitality and modularity. This enables unbounded product-agnostic scale of innovation through combination and recombination of digital resources through the distribution of knowledge and control among heterogeneous actors from different disciplines, communities or stakeholder groups. However, Yoo et al. (2010) acknowledge that different architectural forms would present different organising logics. Within an integral architecture, the dominant organising logic would be a vertically integrated hierarchy (Langlois, 2003; Teece, 1993) wherein a modular architecture involves a vertical disintegration of a firm's design and product functions (Baldwin & Clark, 2000).

Hanseth & Lyytinen (2010) propose to tackle such emerging complexity between the technical architecture and organising processes within its innovation trajectory through a design theory of complex adaptive systems which aimed to address how emerging tensions or paradoxes within the innovation processes are negotiated. Theory on complex adaptive systems draw on complexity theory (Benbya & McKelvey, 2006). Complexity theory is the result of five decades of research into non-linear dynamics across a range of disciplines spanning natural sciences, biology, physics, computer science as well as information systems, resulting in a meta-theoretical framework which can admit diverse epistemologies and methodological approaches (Benbya et al., 2020). Based on underlying assumptions of non-linearity, path-dependency, unpredictability, and emergence this approach sheds light on how systems made up of autonomous agents adapt and evolve as they self-organise in response to changes or stimuli emanating from other agents and the wider system (Holland, 1995; Dooley, 1996). Theory on complex adaptive systems and complexity theory is based on unpacking the reasons behind changes within a system through phase transitions, catastrophic failure and unpredictability. Since complexity theories assume systems to be in a constant state of flux they are premised on the notion that "(t)hey cannot, therefore, be understood by simply examining the properties of a system's components" (Benbya et al., 2020, p. 4). However, in developing their theory of complex adaptive systems of IS innovation, Hanseth & Lyytinen (2010) concur with Yoo et al. (2010) in submitting that architectures that require significant early investment or architectures that exhibit integral characteristics might follow a more centralised specification driven approach. Lyytinen et al. (2016) further suggest that as increased digitalisation leads to increasingly complex forms of product innovation, it requires a nuanced conceptual vocabulary to understand digital architectures and their capabilities for organising the innovation process in relation to heterogenous actors and other artefacts. In explicating its innovation process, this study aims to unpack the nature of the architecture to understand its relationship with the ecosystem it engenders and wider environmental conditions that come to be implicated within the development process.

Complex product innovation involves high user involvement with specialised multi-actor coalitions focusing on given design problems where innovation proceeds through *ex ante* negotiation (Hobday, 1998). Moreover, developing simulations can involve varying assumptions about the nature of the operational domain, selection of parameters and data sources, and modes of validation (Leonardi et al., 2021). Further, the humanitarian sector within which the proposed technology is aimed to be used is a globally interlinked system with multiple actors wherein formal and informal rules within the wider environment shapes the nature of resources and relationships (Rush et al., 2021). The above discussion highlights the importance of the nature of architecture, relationships within the ecosystem, and wider structural conditions that stand in inter-locking relationships within the innovation trajectory.

Synthesising from existing literature on digital product innovation, complex products, and contextualist research with a socio-technical approach, this study contends that the innovation process of complex products is a result of the technical architecture, the generative nature of digital products, and their interrelations with both the proximate and distant context. This highlights the need for understanding and explaining the phenomenon in totality. The imperative for undertaking such a theorisation exercise would involve the joint optimisation of the technical and social and within the social it involves the simultaneous appreciation of structure and agency across multiple levels in relation to the technical architecture.
Given the need for joint understanding of the technical and the social it highlights the four dimensions of socio-technical systems i.e. actor, task, technology, and structure (Leavitt, 1964). The aim is to understand both the ‘engine and content’ of IS innovation in the form of technological elements and functionalities, actors and stakeholders involved in the innovation process, the task or what and how development is achieved, and structural conditions that shape action (Lyytinen & Newman, 2008). This is encapsulated to an extent in structurational approaches to socio-technical studies and Pettigrew’s notion of inner and outer contexts in contextual studies (Avgerou, 2019; Pettigrew, 1997). While the application of these meta-theoretical approaches helps in shaping an understanding of the recursive interdependency between technology and social outcomes (Orlikowski & Robey, 1991), it also becomes important to develop middle-range explanation of IS phenomenon that can provide theoretical insights while straddling the tension between scale and detail or generalisability and relevance (Lyytinen & Newman, 2008).

deReuver et al. (2008) highlight how innovation of digital products is often studied as a snapshot in time that precludes a more holistic understanding of the different dynamics involved in the process. This underscores the need to think of digital product innovation as a socio-technical process (Lyytinen & Newman, 2008). This is because the effect of design choices can only be observed through a longer processual analysis. Design choices can have cascading effects on the evolution of the product and resultant considerations within its ecosystem (Yoo et al., 2010; Ghazawneh & Henfridsson, 2013). This reinforces a need to understand how design choices create antecedent conditions for innovation and product outcomes (Robey & Newman, 1996). Moreover, developing complex products involves a series of choices between different alternatives and negotiation and management of opposing tensions as a result of heterogeneous stakeholders and multidisciplinary teams. This can be further compounded by conflicting objectives with the possibility that the nature of the problem can change over time (Mumford, 2000). Tiwana et al. (2010) highlights the role of endogeneous design choices and exogeneous conditions shaping the evolutionary trajectory of digital products. However, as the preceding discussion highlights, there needs to be a better understanding of the nature of choices and the conditions whereby they come to be translated into product functionalities which are interrelated with their proximate and distant contexts. This further stands in interdependent relationships with the nature of architecture and how conditions shape choices with regard to the technology and how certain conditions about the nature of the technology itself like generativity and digitality enables the exercise of certain choices.

A framework that allows unpacking conditions of choice is Kleine's Choice Framework (Kleine, 2010; 2011). Though borrowed from the allied field of ICT4D it helps unpack underlying choices as a transmutation of conditions of the technical architecture, structure, and agency i.e. conditions across and within layers of context in relation to the technology. In essence, it provides a vehicle for operationalising meta-theoretical imperatives towards developing middle range explanations to understand and explain the process of innovation as an inter-dependent socio-technical relationship that shape design choices and functional outcomes of the product. The subsequent adaptation of the framework for this study helps understand the role of different socio-technical conditions implicated in design choices that result in antecedent conditions for further innovation. The framework builds on the notion of inner and outer context by unpacking the nature and constituents of what enables agency in the form of resources, the conditions within wider the environmental context, and the nature of interdependencies that they jointly create among stakeholders in relation to technological outcomes. Despite its development and application in a different disciplinary area of ICT4D, the Choice Framework provides a way of conceptualising choices by unpacking conditions behind them. This helps determine how choices materialise through *ex ante* negotiations through which complex product innovation proceeds (Hobday, 1998; Hobday et al., 2000). Particularly, in the occasions of high user involvement and dynamic user requirements which in this context involves humanitarian organisations coordinating operations through large globally interlinked systems with significant resource constraints but strong institutional mandates to deliver immediate relief services to populations at risk. This helps understand antecedent conditions of innovation not just as a function of earlier choices but also helps determine the persistence of such conditions or the introduction of newer ones that may continue to shape action and innovation over time. Responding to a call by IS scholars to understand innovation as a process (Fichman et al., 2014) and unpacking conditions underlying choices over time, it helps develop a holistic understanding of digital product innovation through a combined understanding of the technical and social across layers of context. This helps develop explanations based on innovation as a process or series of choices and different combinations of conditions that result in given outcomes or product functionalities. It provides a way for identifying mediating conditions and developing middle-range explanations for the innovation trajectory of complex digital products like ABM while being sensitised to meta-theoretical assumptions of the relationship between structure and agency around technology and action. This would help develop a newer vocabulary for understanding the nature of complex digital product architectures like simulation models and the specificities of their innovation process in the light of wider contextual conditions. The following section outlines the Choice Framework in greater detail focusing on its origins, application, and its adaptation for the contextual study of complex digital product innovation.

3.3 The Choice Framework

The Choice Framework was developed by Kleine (2010; 2011) to understand the role of ICTs in socio-economic development, particularly its implications for development outcomes. It was developed to operationalise Sen's (1999) capability approach to human development which was aimed at understanding how certain conditions or inputs such as technology affects an individual's 'functionings' or their ability to achieve their desired social and economic goals. Its roots lie within the longer legacy of Amartya Sen's work in human development and social choice theory which is invested in understanding the process of how diverse set of conditions within the systemic environment translate to existence and exercise of choice. It is congruous to the notion of the environment as providing rules and resources that shape technology related action (Jones, 1999). Sen's (1977; 1999) work in social choice theory laid the foundation for his later work in building normative and evaluative frameworks of understanding the nature of formal and substantive freedoms in form of individual functioning or what an individual is actually able to achieve given particular conditions and circumstances. As a result, Kleine's Choice Framework is built on normative theories of individual empowerment that helps analyse how certain conditions shape individual abilities to recognise and realise their choices.

The framework emerged in parallel to and was then applied to study the effects of ICTs on microentrepreneurs' livelihoods in rural Chile (Kleine, 2007). Kleine's (2010) Choice Framework is a synthesis and refinement of Alsop & Heinsohn's (2005) work on operationalising Sen's approach and the Sustainable Livelihoods Framework (SLF) developed by the UK's Department for International Development. With the individual as the unit of analysis, Kleine defined components of such a framework to be outcomes, dimensions of choice, agency, and structure where agency is conditional on the resource portfolio available to an individual and structure represents the environment that shapes conditions for a particular choice to materialise in conjunction with available resources. The dimensions of choice involve existence of choice, awareness of choice, use, and achievement of choice and represent the processes which translate choice to outcomes.

3.3.1 Components of the Choice Framework

Agency: Resources form the basis on which individuals exercise their agency to navigate extant social structures (Giddens, 1984). Kleine (2010; 2011) synthesised the two frameworks mentioned above to identify a set of nine resource categories spanning material (e.g. tools or machinery that might be available to an individual), financial (e.g. cash, savings, credit), natural (e.g. climatic conditions, soil fertility, natural resources), geographical (e.g. remote rural or urban), human (e.g. health and education skills), psychological (e.g. self-confidence, tenacity, optimism), information (e.g. local market rates for farmers), cultural (e.g. prestige attached to a professional designation such as academic title), and social (e.g. social connections or group identity). The resource portfolio available to an individual, within their social context, circumscribe the scope of potential action. However, a variety of resources can co-occur unevenly and an individual's agency in achieving their desired goals and objectives is 'inescapably qualified and constrained by the social, political, and economic opportunities that are available' (Sen, 1999, xi-xii). This highlights the simultaneous importance of structure and agency within the Choice Framework.

Structure: The Choice Framework also takes into account structural conditions that enable and constrain agency spanning formal and informal laws, regulations, customs, institutions, policies, programmes, and processes which are highlighted as representative structural conditions. Structural conditions relate in complex ways to the agency element within the framework. For e.g. conditions of access and affordability structure the extent to which an individual with existing ICT skills might be able to use an ICT product. Similarly, in conditions where ubiquitous internet access is not available, in areas such as remote rural ones, an individual can draw on their social resources to access internet at a neighbour's house.

Dimensions of choice: Dimensions of choice includes existence of choice i.e. the existence of different possibilities that can be realised based on the extant resource portfolio and structural conditions, use of choice i.e. whether an individual exercises their choice from the available range of possibilities, and achievement of choice i.e. whether exercising a given choice leads to the achievement of the desired outcome. However, Kleine's fieldwork in Chile led to add another dimension known as sense of choice i.e. awareness of choice which showed that while individuals were aware of the possibilities of new communication technologies like email and online chat, they were not aware of others like Voice-over-IP as a consequence of their level of educational resources and some Chilean media sources stressing the use of some technologies over others.

Outcomes: Individuals use their agency, based on their resource portfolio, to navigate social structures that lead them to desired outcomes. Within the context of the Choice Framework, this is indicated as the observed achievement of desired functionings i.e. 'various things a person may value doing and being' (Kleine, 2011, p. 120) such as being adequately nourished or being healthy or playing an active role in the community. Outcomes acting as the proxy for an individual's actualised choices help map the relationships between resources, agency, and structure that have precipitated within a given observed outcome which in turn help provide a snapshot of an ongoing process.

The Choice Framework has been applied primarily in the field of ICT4D to understand and examine the effectiveness and impacts of digital literacy programmes in Brazil (Poveda, 2016), to explore the use of open educational resources for girls growing up in informal settlements in Nairobi (Zelezny-Green, 2017), as well as evaluating an impact sourcing initiative for people with disabilities (Eskelund et al., 2019). Across each of these cases, the Choice Framework has lent itself to a diversity of methodological approaches. Poveda (2016) used the Framework in a comparative study to evaluate how different pedagogical approaches between two digital literacy courses helped shape learning outcomes. It helped her identify 11 resource categories available to an individual that might increase or decrease based on the pedagogical approach used. Zelezny-Green (2017) used participatory action research to introduce girls to a free app which allowed them to access textbooks and other books on their simple mobile phones. In the ensuing ethnographic work, the Choice Framework helped explore and draw attention to the life situation of the girls, the structural constrains they faced and the resources they could draw upon. Eskelund et al. (2019) developed an interpretive case study using an explanatory mixed method approach of surveys and interviews which were analysed using the Choice Framework which helped highlight the necessity of supporting resources to maximise benefits of government training programmes within extant structural conditions. All of these studies explore the interplay of the structure and agency in relation to the enabling and constraining conditions placed by technology which underlie the theorisation of technology and action as explored earlier in the chapter.

This study draws on the Choice Framework as a contextualising approach to study digital product innovation through the interdependency between the digital architecture, its development ecosystem, and the external environment. However, it might raise the question, despite existing theoretical and methodological richness within contextual IS literature why is there a need for another framework of analysis. Avgerou (2019) highlights two contextualising strategies, the layered and relational each with their ontological assumptions, that helps designate the relevant context. As mentioned earlier, this study acknowledges the role of causal influence across layers of context i.e. the proximate and the distant and their reciprocal interlocking relationship with the digital architecture as well as the structure of relationships within the proximate context. However, identification of the domain of enquiry, designation of the relevant context, and explication of assumptions in the relationship between context and phenomenon requires a framework that explains how successive design choices propel the innovation process in a given way. The Choice Framework helps to unpack multiple conditions implicated in a given choice through its focus on the interdependency between the resource portfolio, interpersonal relationships, and structural conditions. This makes it compatible with pluralistic approaches to causal structure and socio-technical theory of technology used in this thesis. One of the enduring tensions within contextual research is the negotiation between scale and detail, this becomes particularly important when new theory development has to contend with digitality, multilateral inheritances within ecosystems, and explaining reciprocal causal trajectories between digital architectures, their development ecosystem, and their external environment. The Choice Framework helps reconcile tensions across multiple planes of influence to help understand how scale and scope of contextual research for digital product innovation can be negotiated. It helps understand design choices at each stage of digital product development by accounting for reconciliation of reciprocal relationships and conditions of possibility within and across different layers of context in the form of aggregate outcomes that determine each stage of functional extension of the digital product.

Kleine (2010; 2011) developed the framework to understand the role of information and communication technologies (ICTs) within the systemic process of socio-economic development. The Choice Framework was essentially developed to understand the role played by ICTs within the development process. It has been adapted for the study of digital product innovation by synthesising the IS theoretical approach and imperatives within this study to translate its essential elements towards understanding and explaining design choices as opposed to an individual's empowered technological capabilities in use. The framework acknowledges the role of an opportunity structure comprising of a resource portfolio or endowment that is instrumental is shaping an individual's ability to exercise their agency or choice in a particular direction. This in turn is shaped by environmental factors like social, economic, policy, regulatory, and political conditions that are further implicated within the nature and direction of such choice (Kleine, 2010; 2011). Opportunity structure represents the means and conditions which enables an individual or social group to achieve their aims and objectives or the conditions at the intersection of resource driven agency and structure that leads to the materialisation of a given outcome or choice (Latorre-Catalan, 2017). These highlight the processes for understanding conditions and interlinkages underlying successive design choices in relation to the digital architecture.

In conditions of distributed model development with unequal ownership and access to resources, this means expanding the capacity to makes effective choices and translate them into action (Alsop & Heinsohn, 2005). Choice, then, comes to depend on an opportunity structure that individual actors navigate. As a result, it is a function of an individual actor's ability to make a particular choice where choice is dependent on information, organisational, social, financial, human resources among others. Similar to how Alaimo et al., (2020) highlighted the role of data resources in the innovation process, the resource portfolio becomes the 'structure of means' or the basis on which individual stakeholders exercise their choice (p. 43). However, the resource portfolio within the framework involves the existence of a diverse and heterogeneous set of resources spanning material, financial, information, human resources among others. It is the aggregate of actual and potential resources that shape inter-relationships within a group, in this case, an ecosystem. The use of resources as an input by individual actors to exercise their choice can only be realised through their systemic relationship with a given environment in the form of 'formal and informal laws, regulations, norms and customs' (Alsop & Heinsohn, 2005, p. 9). However, this interplay between choice and social structures is also shaped by the interpersonal differences between actors in the form of levels of influence and social conditions which shape the ability to translate resources into choice and action (Sen, 1984). As a result, choice comes to depend on a number of systemic factors that condition the way it can translate resources towards intended action (Kleine, 2011). Structural conditions such as these stand in complex relationships with resource based inputs that shape individual action.

Sen (1999) outlines three ways in which such individual action can be reconciled into collective outcomes. These include the use of an adequate information base (looking at a broader information base beyond individual preferences that take into account interpersonal comparisons), partial resolutions (where full reconciliation between individual interests are not possible), and the nature of relationships (that is dependent on a complex value system based on trust, reliability, etc. beyond restricted self-interest) which shape the actualisation of choice. In the more formalistic branch of social choice theory, collective choice scenarios say in the case of elections, are faced with the difficulty of arriving at collective outcomes on the basis of individual preferences. This is illustrated in the voting paradox: Consider there are 3 options – x, y, and z and 3 individuals 1, 2, and 3. If person 1 prefers x to y and y to z, person 2 prefers y to z and z to x, and person 3 prefers z to x and x to y then majority rule will lead to inconsistencies where x will have majority over y, which will have majority over z unless there is a dictatorial solution of making one person's proposal the overarching choice. Sen (1977; 1999) suggested that a resolution of such deadlock is possible if analysis is decentred from absolute maximisation of individual self-interest. Decision-making often takes into account a variety of factors such as ownership of resources, nature of relationships, conditions of resource provision, and the basis of such resource distribution. As a result, what is possible and what is not i.e. the outcome comes to be dependent on what kind of information is being taken into account. Broadening the scope of analysis by admitting multiplicity of informational bases that are instrumental in shaping social choice help to arrive at consistent and coherent criteria for making decisions when presented with a diversity of conditions. Social arrangements do not require the exhaustive social ordering of all alternative social possibilities. Workable solutions are based on the contingent acceptance of particular provisions and partial agreements without expecting complete social unanimity because insistence on complete social reconciliation hinders practical social action. The basis of action goes beyond self-interest to reflect a range of conditions that form the basis of choice and decision-making. The use of formal economic modes of analyses like the rational choice perspective involve making implicit assumptions to the exclusion of these conditions (Avgerou, 2019). Organisational forms like ecosystems function on the basis of relationships based on norms and values and the amount of power stakeholders exert over the resources that form the components of the system and conditions of their development. Looking at motivations beyond incentive structures based on profit maximisation allows for the variety and variability of conditions that motivate social choices and decision-making. In understanding the role of broad information bases behind action precipitating in a given choice it is important to go beyond the immediate choice of isolated objectives to acknowledge the emergence and endurance of objectives through their effectiveness and survival. Choice can be attributed to direct and derived reasons (Sen, 1999); even if some emerge as the basis of certain choices, their long term survival can structure the evolutionary process of digital product innovation. While the literature in the domain contains implicit assumption of an evolutionary process of innovation, *ex ante* negotiations in successive stages indicate the survivability of unmet objectives which require attention (Hobday, 1998; Hobday et al., 2000).

Application of the Choice Framework would have to contend with the multilateral inheritances within the ecosystem as well as the systemic and structuring effects of the digital architecture and the external environment thereby lending itself to a cross-sectional analysis that is able to establish the linkages between actions and outcomes within a particular opportunity structure. The opportunity structures create the conditions of possibility for successive design choices with the digital architecture, resources, interpersonal relationships and systemic environmental conditions that structure the boundaries within which such choices can be exercised.

3.3.2 The adapted choice framework

Just as Sen (1999) intended his approach to be combined with other theoretical approaches, Kleine (2010) refers to the Choice Framework as a ‘living tool’ to be refined and adapted for different uses. The Choice Framework by Kleine (2010; 2011) was developed to understand conditions of empowerment at the level of use within contexts of socio-economic development. It seeks to understand how the diversity of conditions operating through an opportunity structure translate from existence of choice to their actualisation. IS researchers like Bailey & Leonardi (2015) explore the role of technology choice at the level of use. In their research they ascribe occupational factors over personal preferences as the determinant of particular technology choices within the sphere of work and technology. However, computer scientists and developers also have to make continuous choices with regard to the functionality of the product, the parameters to be included, and datasets and software to be used etc. (Endriss, 2011). The notion of social choice is increasingly being translated to the field of computer science to think about the design and development of computational tools and algorithmic systems. This becomes particularly true in the early stages of the product development where much has to proceed through search and learning and developers have to anticipate and include user requirements (Clark, 1985). Even more so for the development of complex products which operate under high degrees of uncertainty and integral architectures (Hobday, 1998; Ulrich, 1995). The diversity of applications of the approach across domains and contexts of application highlight the versatility and flexibility of this approach to develop holistic explanations in wide ranging disciplinary areas. The Choice Framework is adapted for the study of digital product innovation to understand how the intertwinement of digital technology across contexts shapes the innovation trajectory.

The adapted choice framework aims to understand how interdependencies between the architecture and diverse conditions in proximate and distant contexts shape endogenous design choice in successive stages of development. It begins by developing thick descriptions of the nature of the architecture, ecosystem, and environment through the data collected. This becomes the basis for developing causal claims based on theoretical sensitisation of the interdependency between the three core constructs for explaining the phenomenon of complex digital product innovation and represents the first building block in moving towards theorisation. The components of the adapted choice framework described below as they relate to these three core constructs highlight the mediating conditions implicated in bringing about change at a point in time as well as across time. These become the basis of assumptions within the causal structure about causal influences on the architecture, ecosystem, and environment and thereby, the second important building block for analysis and theorisation after thick causal concepts. The key components of the Choice Framework include outcome, agency, structure, and dimensions of choice (Kleine, 2010). From the literature we know complex products proceed though temporary multi-actor alliances which means a staged phase-wise development (Hobday, 1998; Hobday et al., 2000). This necessitates the need to understand digital product innovation as a process of successive outcomes. The full spectrum of dimensions of choice are difficult to capture or pin down systematically and become obvious when they have led to an observable functional development (Kleine, 2011). Therefore, outcomes act as the proxy for choices that have been actualised, achieved, or exercised. Like the original, the adapted framework starts from outcomes but includes successive outcomes that help determine upward causation propelling innovation (Winter et al., 2014; Pettigrew, 1997). Within the digital product innovation process each phase represents an outcome, where the outcome is congruent to a given choice of functionality towards a design problem, which corresponds to a given resource portfolio, actor configurations, and environmental conditions. Successive phase-wise configurations coalesce to form the direct and derived conditions driving innovation. The sum of all phases helps to arrive at descriptions of the architecture, ecosystem, and environment. These descriptions help illustrate the domain of enquiry and relevant context becoming the basis for the theorisation of the relationship between the contextual conditions and the phenomenon.

The original framework is premised on individual’s capabilities to make empowered choices, the adapted framework take into account stakeholder entities within the ecosystem as individual actors. This is because establishing causal chains from the level of individuals might be long, complicated, and infeasible in explaining a given phenomenon. Looking at collective actors are therefore an accepted strategy to cope with theoretical complexity and aligns with the cross-level contextual analysis used in this research (Markus & Robey, 1988; Avgerou, 2013). Agency of individual actors, particularly the system integrator comes to depend on the resource portfolio. Within the socio-technical view and contextualist approach, agency is constrained not just by the social, economic, and political conditions but also conditions within the technical architecture and multiple planes of influence that cuts across layers of context - in other words the interplay between actor, structure, task and technology (Sen, 1999; McLeod & Doolin, 2012; Kleine, 2010; 2011).

As highlighted above the application of Choice Framework and its adaptations have yielded identification of newer resources in the resource portfolio, methodological diversity, as well as the interplay of structure and agency wherein the technological artefact is situated.

Therefore, the adapted choice framework helps unpack the interdependency between the core constructs of the literature in the following ways:

Choices, outcomes, and architecture: In the original framework, outcomes refer to observed achievement of 'functionings'. As highlighted above, deploying the choice framework for the study of complex digital products like ABM models would need the identification and unpacking of successive outcomes i.e. functional extensions or development of the architecture in successive phases within the innovation process. Moreover, as the full spectrum of dimensions of choices are difficult to capture or pin down systematically, outcomes become the proxy for existing choices that are actualised. Unlike the original, collective actor groups rather than individual persons are taken into consideration because aspects such as awareness of choice entail long causal chains that are difficult to determine. As a result, functional extensions of the architecture are taken to be outcomes of design choices that have been actualised, achieved or exercised. Outcomes become the starting point for socio-technical enquiry to unpack the role of different resources, actor configurations, and environmental conditions that jointly shape both a given outcome and the process as a whole within the opportunities and constraints of the technical architecture and how it organises the innovation process.

Agency, interpersonal relationships, and ecosystem: In the original framework resources become the conditions whereby individuals exercise their agency. Further, the interplay between outcomes and social structures come to depend upon the interpersonal differences between actors in terms of their levels of influence. Within the adapted framework, too, resources form the input on the basis of which collective actors undertake particular actions. The resource portfolio becomes the 'structure of means', whereby certain actions are undertaken which, in turn, enable the system integrator to undertake particular design choices with respect to the architecture. While Alaimo et al. (2020) have highlighted data resources as an important aspect of innovation, other resource categories need to be identified. This is because just as the Choice Framework, invested in the evaluation of ICT related development outcomes, proceeds from an understanding of the development process, its adapted version too must proceed from the understanding of the process of digital product innovation and identify the resource categories implicated within such a process. With different resources co-occurring unevenly, the exercise of agency based on such resources come to be circumscribed by structural conditions and the material capabilities and actual demands of the technology in question (McLeod & Doolin, 2012; Leavitt, 1964). With the architecture organising innovation among heterogeneous actors in an ecosystem, the interpersonal relationships within the proximate context become the mode whereby the structure of means created by the resource portfolio, technical and environmental conditions are negotiated.

Structure and environmental conditions: Agency or the ability to achieve desired goals and objectives is 'inescapably constrained' by the structural conditions in the form of social, economic and political conditions that enable or constrain agency (Sen, 1999). Within this study this translates to the outer context of the environing domain that creates the conditions of possibility precipitating in a particular phenomenon (Pettigrew, 1997; Avgerou, 2019). As ecosystems are composed of heterogeneous stakeholders who are involved in the innovation process through *ex ante* negotiations, they derive purpose, meaning and structure from the contexts in which they are embedded creating multilateral recursive inheritances from different contexts (Winter et al., 2014; Hobday, 1998). The structuring conditions of the environment do not operate in a deterministic way but form an order of rules and resources that shape the feasibility and appropriateness of action within a given context which then comes to be circumscribed by the opportunities and limitations posed by the technical architecture (Jones, 1999).

Once the two building blocks of thick description and mediating conditions are identified, the innovation trajectory of complex digital products is analysed using insights from Sen (1999). According to Sen (1999), the realisation of a given choice through the negotiation of these complex relationships occurs through broadening of information bases, partial resolutions, and nature of relationships. The adapted framework considers the broadening of information bases to be congruous to the scope of contextual conditions i.e. the variety of contextual conditions present within a given phase for a particular choice to materialise. Partial resolutions have been an important aspect of technology development, particularly complex products, which proceed through *ex ante* negotiations, partial attainment of goals or privileging of some goals over others (Hobday, 1998; Hobday et al., 2000; Heracleous & Barrett, 2001; Sabherwal & Newman, 2003). A socio-technical approach to theorising admits the wide diversity of relationships beyond economic and managerial ones. Therefore, scope of contextual conditions, partial resolutions, and non-economistic considerations becomes first component of analysis to understand how mediating conditions transmute into given outcomes. Consequently, looking at successive phases cumulatively helps arrive at the direct and derived conditions of the causal trajectory of the complex digital product innovation thereby helping to move from empirical statements to theoretical statements. As a systemic process based approach, the Choice Framework provides a mode of unpacking the negotiation between different socio-technical considerations within a given phenomenon (Kleine, 2011). The adapted framework helps to understand how such diversity of conditions and complex intersecting relationships translate to digital product innovation. The adapted choice framework shares the limitations of the original framework. Its holistic orientation in theorising the relationship between the key constructs of the architecture, ecosystem, and environment through outcomes, agency, and structural conditions, entails a trade-off with the in-depth theorisation of each element where each of them are products of extensive theoretical traditions with diversity of debates around the comparative influence of one over the other (Kleine, 2010). Further, since it is analytically reliant on observable outcomes as a starting point, it cannot identify negative outcomes or occasions of failure or conditions where choices could not be actualised. Figure 1 represents a simplification of the theoretical approach to unpacking socio-technical interdependencies within complex digital product innovation.

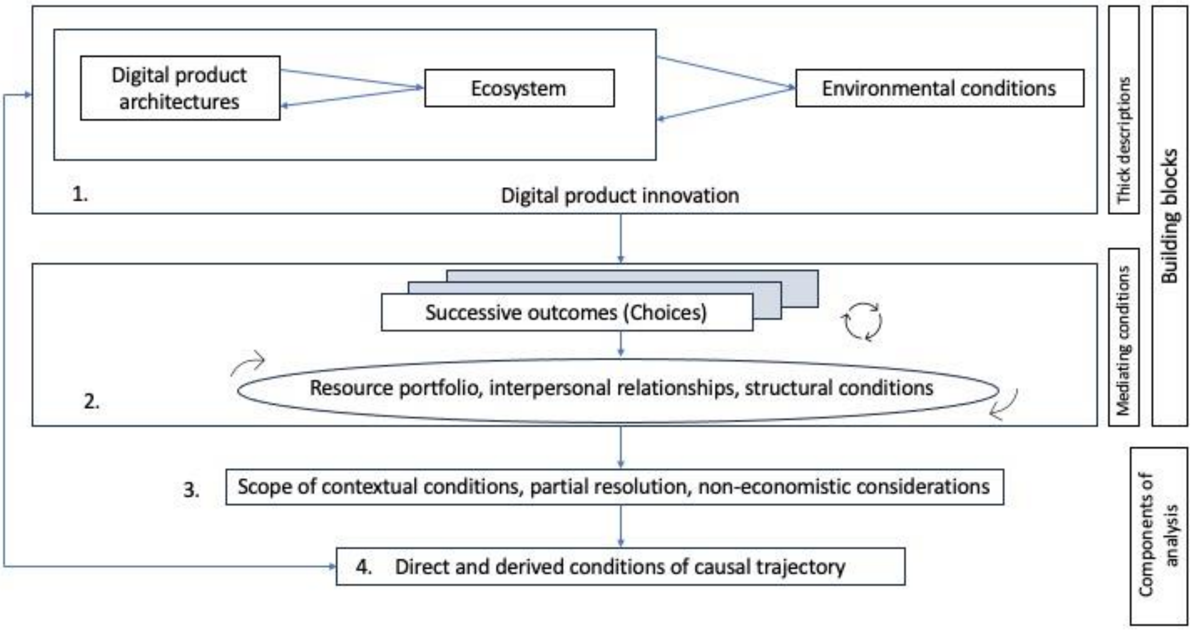


Fig. 1: Contextualising digital product innovation through adapted choice framework

As mentioned earlier, the literature on digital product innovation helped highlight the interdependency between the digital architecture and ecosystem wherein the architecture organised heterogeneous actors within the innovation process and innovation proceeded through negotiation of diverse stakeholder interests. The literature on complex products and contextual studies in IS highlighted the role of the external environment. Therefore, architecture, ecosystem, and environment emerged as core constructs in theorising the phenomenon of digital product innovation through their interdependent relationships. Unpacking the nature of interdependency would shed light on the causal trajectory of the phenomenon i.e. how do multiplicity of conditions translate and evolve into particular outcomes within this process. This interdependent relationship is unpacked through application of the adapted choice framework wherein successive outcomes are examined through the mediating conditions of resources, interpersonal relationships, and structural conditions which help unpack the conditions of agency and structure in relation to the technology or in other words identification of conditions of resource availability and structural constraints that get negotiated through interpersonal relationships into observed outcomes. This entails developing thick descriptions of the nature of the architecture, ecosystem, and environment (Layer 1 of Figure 1) and identification of the mediating conditions of resources, relationships, and structural conditions that translate into particular product outcomes (Layer 2 of Figure 1) which serve as building blocks for theorisation. Building on a longer legacy of Sen’s work with insights from contextual research, their role in the causal trajectory is unpacked by identifying the scope of contextual conditions in each phase, the role of partial resolutions, and non-economistic considerations that translate into given outcomes (Layer 3 of Figure 1). Looking at successive outcomes cumulatively helps determine how both isolated objectives implicated within a given outcome and the survivability of unmet objectives shape the innovation trajectory of the digital product within its proximate and distant context of the ecosystem and environment respectively (Layer 4 of Figure 1). This, in turn, helps develop explanations for the innovation of complex digital products through theorisation of the nature of interdependencies between the architecture, ecosystem, and environment.

4. METHODOLOGY

Chapter summary: *This chapter outlines the research design as a single embedded case study and examines its suitability for the proposed direction of theory development. Key aspects of the research design such as case selection, unit of analysis and data organisation are explored. The data is organised as process research acknowledging circular and reciprocal causal relationships instead of linear cause-effect ones using a priori constructs from literature and theoretical sensitising devices to develop explanations for the phenomenon under consideration. This chapter also contains a first person narrative account of doing field study driven process research, the complexity of managing and organising process data, and operationalising the adapted choice framework for analysis.*

4.1 Research design

A research design is the logical sequence that drives a study from observation to theorisation. It helps take the study from a research question to a conclusion through empirical data (Yin, 2014) thereby determining the journey through which causal explanations underlying the process of complex digital product innovation can be derived. This sequence involves a number of steps including collection, analysis, and interpretation of the data (Nachmias & Nachmias, 1992). This research is designed as a single embedded case study organised as process research. This is in line with the adapted choice framework in viewing complex digital product innovation as a process of successive stages each corresponding with a design choice and corresponding outcome that advances product development. As empirical methods, case studies help explore a phenomenon in its real world context where important contextual conditions are likely to play a significant role (Yin, 2014; Yin & Davis, 2007). Therefore, it helps in detailed exploration of the socio-technical interdependencies underlying the phenomenon under observation. As a research strategy, case studies aim to understand the dynamics of a phenomenon in particular settings (Eisenhardt, 1989). Case studies are a widely used research method within IS providing the opportunity to get detailed information about a particular phenomenon (Eisenhardt, 1989). While single embedded case studies allow for rich description and acknowledgement of wider contexts, it poses limitations due to the lack of comparability which could undermine causal explanations drawn from the data and constrain generalisation of findings. However, this limitation can be partially overcome by applying case methods over a longer period of time which allows for the observation of a range of different types of events and helps to distinguish significant events from ordinary ones in moving towards theoretical generalisation (Lee & Baskerville, 2003). Since case studies stress on developmental factors i.e. a case typically evolves over time as a result of specific and interrelated elements which constitute the case as a whole (Flyvbjerg, 2011), it is particularly well suited for process research. This approach is particularly suitable for the study of complex digital product innovation since it helps understand the entity that emerges from multiple planes of influence (Mills et al., 2010). By helping to derive explanations for complex causality, they enable multi-level analysis since outcomes can be a result of multiplicity of conditions underlying choices driving digital product innovation. Being particularly suited for answering how and why research questions, it helps to link outcomes to the conditions that help bring them about (Yin, 2014; Avgerou, 2013).

The main components of a research design include: the research question, unit(s) of analysis, and the criteria for interpreting findings (Yin, 2014). Theory building from case study research involves a journey from specifying research questions to reaching closure. As a result, case studies benefit from prior development of tentative *a priori* constructs to guide design, data collection, and analysis (Eisenhardt, 1989). *A priori* constructs can guide the researcher's attention towards particular areas of interest which helps in managing research objectives and data collection and interpretation. The literature review and the theory section helped to identify tentative *a priori* concepts that act as sensitising devices to help focus attention on particular aspects that should be examined within the scope of the study as synthesised in Figure 1 in the last chapter (Eisenhardt, 1989; Yin, 2009; Klein & Myers, 1999). This research works with digital architectures, ecosystems, and environment as key constructs within digital product innovation as identified through literature. Contextualised socio-technical interdependencies between them acts as sensitising devices to guide theory development through explanations derived from implementing the adapted choice framework. The following sections highlight the rationale for case selection and explication of unit of analysis, followed by the organisation of data through process research, the role of explanatory research, and the journey from observation to theorisation through generalisation and abstraction.

4.1.1 Case selection and unit of analysis

Case selection is one of the important aspects of case study research as it circumscribes what can be learnt and contributed (Flyvbjerg, 2011). Case selection often depends on the underlying motivation for research i.e. theory testing or theory building. In the former a pre-existing relationship identified from theory is used as the basis for formulating hypotheses which is then statistically tested. In qualitative causal process research approach within IS, as used within this study, theory testing and theory building cannot be separated from each other (Avgerou, 2013). IS theorisation is rarely purely inductive and is guided by *a priori* constructs from the literature (Avgerou, 2019). Case selection can be done in two ways: through random sampling or through information-oriented approaches. Random sampling involves the selection of a representative sample to avoid systematic biases for the purpose of generalisation and is used in hypotheses testing for theory-testing studies. In an information-oriented approach cases are purposively selected on the basis of the information they can provide about a theoretical area of interest (Flyvbjerg, 2011). They are more commonly used in theory building exercises to develop qualitative causal explanations in social theory driven IS research. Cases may be chosen to replicate or extend emergent theory or fill theoretical categories by providing polar examples (Eisenhardt, 1989). Case selection is therefore driven by a choice that can provide useful variation on the dimensions of theoretical interest (Seawright & Gerring, 2008). Pettigrew (1988) suggests that given the limited number of cases that can be studied, it is judicious to choose cases representing extreme situations and polar types where dimensions of interest are transparently observable.

This research employed an information-oriented case selection strategy and theoretical sampling. The study of the Flee ABM developed by the BUL's Migration Modelling and Simulation Group (the team) provides an opportunity to make a theoretical contribution by studying the innovation process of complex digital products like ABM which has largely been underappreciated in existing literature. However, one of the unacknowledged aspects of case selection involves question of access i.e. whether access yields the richness of information through which insights can be built for theory development. Considerations of access do not just include access to a physical site or someone's time but can be multidimensional in terms of social conditions that structure the information provided and obtained (Fjellström & Guttormsen, 2015). Studying the innovation process of complex digital products is inevitably complex in itself given the diversity of technological approaches and computational techniques involved. Not least when such specialisation is distributed among multiple projects and collaborators as was in the case of Flee. Starting with the initial motivation of studying how emerging technologies are developed, there were three possible candidates cases: one in the development sector involving an AI solution developed by a research institution in a developing country for the implementation and management of government health programmes in a particular region; another was an integrated AI and drone capability developed by a multilateral humanitarian organisation for post-disaster damage assessment; and the third was the Flee ABM case. All three cases were in the social sector to ensure ease of access and freedom and openness in the research process. Among the three cases, the BUL team being an academic group understood and provided the full and unfettered access required for an in-depth academic enquiry including often sitting down to explain difficult technical terms and approaches. While the other two were promising solutions, COVID-19 restrictions were in place during the time when this research was conducted which prevented co-location in those organisations. This constraint meant relevant individuals within those organisations had to specially carve out time to talk to the researcher through video-conferencing complicated also by not being present while particular product decisions were taken. BUL being located in London enabled in-person participation when COVID-19 restrictions were relaxed. Further, since the team already worked with project-based distributed teams across different locations, many of these meetings were already remote first which helped to observe the different aspects of the product development process as they unfolded.

Case studies can be single or multiple and each of them can involve single or multiple units of analysis (Yin, 2014). While a single case study approach privileges the logic of theoretical sampling which helps to extend and / or augment existing theory, multiple case studies are based on the logic of replication which can be either literal (i.e. case selection is predicated on expected similar results across cases) or theoretical (where cases present contrasting results for expected reasons). Yin (2014) highlights five rationales for selecting single case studies: critical, unusual, common, revelatory, or longitudinal. Given that theory is the substantive context behind each of these rationales, a critical case starts with clearly identifiable propositions from theory and provides the opportunity for significant contribution by confirming, challenging, or extending such theory. A critical case is one which has strategic importance with regard to the general problem offering the possibility for generalisation in the form of “if it is valid for this case, it is valid for all (or many) cases” or “if it is not valid for this case, then it is not valid for any (or only few) cases” (Flyvbjerg, 2011, p. 307). An unusual or extreme case is one that deviates from theoretical norms, thereby offering a unique opportunity for documentation and analysis which might reveal greater insights about normal processes. Extreme cases help reveal the limitations of existing theory and put forward concepts, variables, and theoretical contributions about phenomenon previously considered an outlier (Flyvbjerg, 2011). A common case is one that highlights a particular relationship previously identified in theory and may also be called a representative case. A common case is often well explained in the literature, hence the research problem is located within the case. Common cases are often selected when the researcher wants to identify alternative causal mechanisms within a known phenomenon (Seawright & Gerring, 2008). A revelatory case is contingent on the opportunity to study a phenomenon that was previously not accessible for social enquiry. A longitudinal case is used when changes over time are of interest and where the theoretical dimension involves an enquiry into how conditions and processes change in relation to time. While different theoretical objectives may underlie the rationales for case selection, they are not necessarily mutually exclusive; for example, a case can be both extreme and critical. The types of knowledge claims made from the case depends on how the case is viewed in relation to theory (Flyvbjerg, 2011). ABM is an unusual case because it offers distinct variation from the paradigmatic cases of modular architecture on which digital product innovation theory is premised. However, it is also a common case of digital products since they find diverse range of applications and are also representative of classes of digital products characterised by specialised high technology developed in non-mass market conditions before being commoditised like machine learning, large language models, and image classification algorithms etc. It is longitudinal in the sense that it traces the digital product innovation process through historical reflection of the design and development of the product.

Selection of case studies sets boundaries around the extent of knowledge claims, as a result, selection and bounding of the units of analysis within a particular study are significant (Ragin, 1992). Case studies often focus on drawing the boundaries around individual units of analysis thereby helping separate the case and the context. Single case studies can either be holistic or embedded. Holistic case studies explore the global nature of the entity or phenomenon under consideration. This involves a single unit of analysis where the case and unit of analysis are on the same level. An embedded case study involves more than one unit of analysis within a single case. A holistic case design is better suited when no logical subunits of analysis can be identified or the underlying theory itself is of a holistic nature (Yin, 2014). However, holistic case studies operate at a higher level of abstraction and inhibit a researcher from studying the case in operational detail or providing a nuanced understanding. With an embedded design, while a single case can have more than one unit of analysis, the individual analyses of each of these units must return to the primary unit of study in order to make a contribution. Additional units of study provide a significant opportunity for nuanced and extensive analysis thereby enhancing contribution to theory. Tiwana et al. (2010) highlight the importance of being explicit about the unit of analysis as the nature of insights depend upon the unit from which observations are derived. While the primary research question engages with the socio-technical interdependencies underlying digital product innovation, the operative research questions seek to unravel these interdependent relationships with the digital architecture by focusing on the ecosystem and the environment respectively. These embedded units of analysis help in bringing the focus back to the primary research question in answering how both proximate and distant conditions, in conjunction with the digital architecture, shape the focal phenomenon of complex digital product innovation.

4.1.2 Organising data: process research

Process has been used in three ways in the literature: (1) as logic used to explain causality in variance theory; (2) category of concepts referring to organisations and individuals; (3) sequence of events to describe changes over time (Van de Ven, 1992). Pettigrew (1997) argues that out of these only the third is able to observe process in action and is able to account for how a phenomenon develops and changes over time. As highlighted in the last chapter, the development of complex products can take the form of successive multi-actor alliances which necessitates the need to look at digital product innovation as process or sequences of events or phases. Pettigrew defines process as “a sequence of individual and collective events, actions, and activities unfolding over time in context” (1997, p. 338). Process studies are deployed for describing, analysing, and explaining the what, why, and how of some sequence of individual and collective action. The underlying assumption within process studies involves a dynamic view of reality with processual analysis attempting to “catch reality in flight” (p. 338). While actions drive processes, actions are embedded in contexts which limit their information, insight, and influence but action and context exhibit a dual influence of mutual shaping. Processual analysis aims to understand the cumulative interchange of actors and contexts over time. Therefore, past events shape ongoing and future events where explanation for a particular event depends upon its locations within the processual sequence. While time and history are crucial for processual analysis, it does not stop with a historical account of a given case but uses it for repetitive questioning to look for underlying conditions that drive a process. These underlying conditions can be directly observable as a part of conscious intentions of actors or a part of the immediate or distant context. As a result, process research involves cycles of deduction and induction (Pettigrew, 1997). Deduction involves foresight about primary purpose of the research, themes, and questions. This arises from an evaluation of existing theory which comprises the conceptual vocabulary of the research. The deductive approach provides a segue into more open ended inductive exploration that involves constant engagement with the data for continuous reasoning and pattern recognition. This marks the reciprocity of the deductive-inductive cycle where findings are compared against the existing theory and theory development in turn progresses through this continuous mutually informing cycle.

Case studies can involve a multiplicity of data collection methods like archives, interviews, observations etc. and can be used to accomplish various aims like providing a description, testing or generating theory (Eisenhardt, 1989). Process data can be messy, difficult to organise and manipulate with multiple levels and units of analysis: it can consist of events, activities, and choices ordered over time with their analysis involving conceptualising events and detecting patterns (Langley, 1999). While organisation of process data can take many forms the most common pattern involves a sequence of ‘phases’ that occur over time. Attending to context is one of the principal reasons for adopting a qualitative process approach (Pettigrew, 1992; Yin, 2009) which invariably leads to multiple levels of analysis that are often difficult to separate from one another. Despite the focus on events within process research, events can mean anything ranging from “a bad year, a merger, a decision, a meeting, a conversation, or a handshake” (Langley, 1999, p. 693). While the historical data within process research may be sparse highlighting only memorable moments and broad trends, more recent data will be fine grained and will require substantial distancing to separate out significant aspects from “mere noise” (p. 693). As a result, process data can be eclectic, and can pose considerable challenges. The complex nature of process data is representative of the complex nature of the social reality that it tries to capture. This highlights a need to organise data in a way that acknowledge the presence of multi-layered contexts, multi-directional causality, and feedback loops. The primary data in this research consists of field notes from non-participant observation, academic publications of the team's research findings, project reports of deliverables as well as internal reports or working documents. It also involved preliminary interviews to get an overview of the product, its history and current direction which served as orientation when beginning the field work. Given the multi-context analysis, this also involved identifying key environmental conditions through such observations and documents followed by their independent investigation. The documents and observations were used to construct a timeline of phases. Each phase represented an outcome serving as a proxy for choice wherein underlying conditions of the digital architecture, ecosystem, and environment was explored. Within each phase, actor configurations, project affiliations, and resources were identified through co-authorships and acknowledgement sections, particularly for phases preceding the observation period. The environmental conditions beyond actor considerations were identified through observations, informal chats, and references in documents. This also included the organisation of a multistakeholder meeting supported by the LSE Knowledge Exchange and Impact fund in partnership with the BUL team. This brought together key humanitarian organisations and technical actors to discuss issues around emerging technologies in the domain. This helped highlight issues at the intersection of technology, organisational capacity, and wider context. Construction of process timelines as presented in the Appendices 1 and 2 required definitional work on resource categories and environmental drivers through multiple engagements with the data to ensure its effective organisation for analytical purposes. The timeline so constructed was presented to the BUL team lead to ensure correctness. The first review included presenting a preliminary version of the process timeline to the team to ensure accuracy of the flow and constituent elements. The second review included sending the Chapter 5 and Appendices 1 and 2 to the team lead. This included explanation of definitional categories and how they were derived. The reviews elicited supportive examples, minor corrections, and insights which were then incorporated. Minor corrections have been directly incorporated in the text while additional insights have been attributed to the review process. The detailed outline of each phase helped establish the overall development process and work out descriptive aspects of social and technical elements of the domain of enquiry, the relevant contexts, the resource portfolio, and environmental conditions by applying the adapted choice framework at each phase. This then formed the basis of deriving explanations for upward causation from successive phases to the overall process of complex digital product innovation.

Structuring data into phases or stages helps in simplification of a dynamic reality and becomes the first step in establishing the circular and reciprocal causal relationships within social theory driven IS research (Pettigrew, 1997; Avgerou, 2013; Orlikowski & Baroudi, 1991). This helps processual analysis work towards holistic explanations by broadening the field of analysis from nested contexts to wider conditions of influence and possibility (Avgerou & Madon, 2004). Process analysis aims to explain the links between context, processes, and outcomes which involves studying processes across layers of context and incorporating historical reflection through a holistic rather than linear view of process (Pettigrew, 1997). This helps acknowledge the relationship between socio-technical interdependencies and action in explaining outcomes. This is because process comes to be shaped by inner and outer contexts i.e. both the immediate setting of the phenomenon as well as the broader environmental conditions. This highlights the need for a multi-level analysis by linking outcomes to different levels of the relevant context, in this case the proximate context of the ecosystem and the distant context of the environment. However, the levels of context to bring into the research depends on the content of the research problem, research themes, and questions driving the study. Further, it becomes important to understand the sequences and flow of events over time to unpack antecedent conditions shaping outcomes (Robey & Newman, 1996). It becomes necessary to look beyond surface events and chronology to search for patterns and structure that help explain context and action (Pentland, 1999). Time and historical reflection involve judgements about when the process begins and ends and the selection of this time period becomes the organising mechanism for analysis. Context is not just the stimulus environment but a nested arrangement of structures and processes which actors negotiate (Pettigrew, 1997). This reciprocal relationship between context and action view context and structure not just as barriers to action but as essentially involved in its production showing how aspects of contexts can be mobilised by actors towards desired outcomes. Causal linkages within the context can only be exposed by understanding the recurring patterns over a period of time. In holistic explanations, context is neither linear nor singular and process analysis involves looking for proximate not final causes i.e. the multiple intersecting conditions that link context and process to outcomes. Acknowledging the multi-level, multi-actor, historical process involves utilising research design to reduce the complexity i.e. to have a clear outcome to begin process research.

Process narrative is one of the strategies to organise and manage this data in a meaningful way that is able to convey the dynamic richness of the given social reality but at the same time is understandable and potentially useful to others (Pentland, 1999). The following chapter constructs a narrative based on phase-wide identification of conditions underlying choices at each stage. It begins with an overview of the process followed by unpacking the multiplicity of conditions that helped in constructing descriptions for each element of theoretical interest i.e. digital architecture, ecosystem, and environment. These descriptive accounts were particularly important in understanding how particular architectural forms organise innovation and how the innovation process is managed within the resultant ecosystem in the light of wider environmental conditions. The narrative strategy is a favoured approach within contextualist research since it helps manage the complexity of the data and clarify sequences and causal linkages across levels of analysis (Pettigrew, 1990). Because of the level of contextual detail, this approach is well suited to single case studies. The variety and richness of detail within the narrative should be able to convey a degree of authenticity that might not be possible to glean from large samples (Golden-Biddle & Locke, 1993). However, when using the narrative strategy, one must be mindful of moving beyond an idiosyncratic story of the phenomenon to offer more explicit theoretical interpretations. The adapted choice framework is implemented for each phase which corresponds to a given outcome and choice. Resource portfolio, nature of relationships, and environmental conditions underlying each phase highlights the persistence or dispensation of such conditions over time. This helps trace upward causation from single events towards a causal trajectory of digital product innovation to understand how it is shaped through successive socio-technical configurations over time as they present direct conditions within a given phase and serve as derived and antecedent conditions for successive phases within the innovations process as a whole.

Narratives contain events occurring in a sequence in time which forms a part of the deep structure of the phenomenon. Even though it is not necessary for the surface structure of the narrative to present events in sequence, chronology is an organising logic (Pentland, 1999). However, restricting process narrative to sequential events only runs the risk of “narrative positivism” as it systematically excludes features needed to construct explanations beyond descriptive accounts (p. 713). This highlights the need to delineate the domain of enquiry and the relevant context which help clarify the level of analysis. Narrative structure then comes to include the nature of links and relationships between social and technical elements. Constructing explanations linking outcomes to actions and conditions require an enquiry across contexts to explicate the way multiplicity of interests and conditions are negotiated through particular actions on to particular outcomes. Attention to dimensions that transmute conditions and negotiate interests into action can be helpful to uncover unstated assumptions underlying seemingly managerial and technical rationality.

Process research resolves a conflict between better stories and better constructs i.e. sacrificing context and deep structure to build better constructs and vice versa (Pentland, 1999). This is because narratives help explain the relationship between different events within a process while containing insights for an underlying construct. Within process research, observed patterns of events become core theoretical explanations. Narratives become abstract conceptual models used to explain observed data by helping circumscribe a wider aspect of the observed phenomenon. One of the critical challenges is to move from surface to deep structure with the data that is collected representing the surface structure of the event. Another involves crafting a single account out of partial, subjective, and even conflicting accounts. Moving from surface observations to underlying structures involves a move from description to explanations (Simon, 1992). Process theory helps highlight the conditions for change and evolution. Change occurs at multiple levels with the coalescing of conditions into a larger innovation process.

(a) Doing process research

Theory development through methodology requires the specification of the researcher's ontological and epistemological stance since they frame assumptions about ways of knowing (Avgerou, 2013). While located within an interpretive approach, it more closely adheres to Avgerou's (2013) position of tracing causal explanation within social theory driven IS research. In essence, this avoids the strand of interpretive research that espouses a completely social constructionist and anti-causal position that validates description as the only form of possible knowledge claim (Martin, 2011). It adheres more closely to a pragmatist worldview where thick descriptions from an interpretive study help derive theoretically generalisable conceptual explanation through analysis (Avgerou, 2013; Gross, 2009; Lee & Baskerville, 2003). Working within a pluralist approach to causal structure as identified in the last chapter, it becomes necessary to avoid the extremes of universalism (explicating an unknowable objective reality which exerts causal effects) and relativism (in the form of radical social constructionism). The pragmatic approach holds that means and ends are not given prior to action but comes to be constituted either through past learning formed by historical social conditions or by initiating new lines of activity when newer problems, situation, or events present themselves (Gross, 2009). This helps in understanding how socio-technical configurations change or stabilise over time through influence from multiple contextual levels to sustain and propel the innovation process. The adapted choice framework helps unpack these instances across phases by identifying the strands of circular or reciprocal causal relationships that are usually uncovered through interpretive research (Orlikowski & Baroudi, 1991).

Beyond the sequence of events, process narratives involve identifying explanation or linkages between conditions and action. These have been called social mechanisms (Avgerou, 2013), generating mechanisms (Pentland, 1999), and motors (Van de Ven & Poole, 1995). They help explain how a set of entities and activities produce change from an initial state to observed outcomes (Gross, 2009; Hedström, 2005; Hedström & Swedberg, 1998). They form the intermediary level of analysis between description, generalisation, abstraction, and theorising (Hedström & Swedberg, 1998) and provide the building blocks for the construction of social phenomenon (Van de Ven & Poole, 1998). Explicit identification of such mechanisms, motors, or causal claims helps make "the constituent parts of causality surface to form explanatory theory of complex social phenomena" (Avgerou, 2013, p. 407). It leads to an approach to theory development involving multi-causal explanation in IS research.

Unlike variance research where such causal claims are identified through theory to formulate hypotheses that are then tested statistically, a social theory driven process approach in IS shows no linear cause-effect relationship (Avgerou, 2013). The explanation that is constructed by process tracing takes an approach to theory development that while being informed by *a priori* constructs is able to contribute to theory by identifying and explaining phenomenon occurring in different conditions and contexts. Further, unlike variance models which explain the existence of a relationship between the dependent and an explanatory variable, process narratives help reveal the logical link between conditions and outcomes by tracing causes in sequence of actions and events that connect them. While variance models aim to establish uni-directional cause-effect relationship, process narratives suggest circular and reciprocally interacting approaches to causality (Markus & Robey, 1988; Orlikowski & Baroudi, 1991). A major challenge in identifying causal claims is the tension between identifying something specific enough to have explanatory value for a particular observed phenomenon while at the same time general enough to apply to different empirical fields.

Therefore, it becomes important to derive portable concepts from the specificity of particular cases that can operate in different contexts without specifying outcomes (Falleti & Lynch, 2009). However, causal explanations derived from such research do not have predictive value, they are indeterminate and generalisation falls in the category of deriving abstractions from the specifics of a case (Lee & Baskerville, 2003; Eisenhardt, 1989).

Causal claims operating at organisational and social levels of analysis are traced within narrative accounts of processes (Eisenhardt, 1989; Pentland, 1999). However, narrative accounts describing sequence of events do not always amount to identification of causal explanations (Avgerou, 2013). Identifying causal conditions that drive processes can be challenging because it might reveal multiple layers of processes embedded in multiple layers of context (Pettigrew, 1997). However, identification of such causal claims help elicit the reasons and conditions that drove actors to act in particular ways.

(b) Developing explanations

Contextual explanations involve important assumptions about the nature of causality (Avgerou, 2019). Explanation requires unravelling causal processes about a theorised phenomenon by helping answer why or how a phenomenon occurs (Avgerou, 2013; Markus & Robey, 1988; Pentland 1999; Van de Ven & Poole, 1995). Development of explanations are linked to development of theory where such development draws from theoretical assumptions about technology and human action (Avgerou, 2013). Such assumptions act as sensitising devices around the choice of elements and relationships through which explanations about a phenomenon are constructed (Garfinkel, 1981; Klein & Myers, 1999). An important aspect of identifying the object of study within IS research is shaped by the theoretical approach to technology and action adopted by the researcher. Such theory guides the selection of some focal entities over others including the meaning, associations and constructs that are ascribed to it. However, IS research often involves complex choices about the selection of appropriate theories of technology and action. “Theorists draw from a continuum of highly abstract foundational theories that explain the world we live in, theories of bounded generalization that apply to phenomena deemed similar, and more narrowly bounded, phenomenon-specific theories” (Avgerou, 2013, p. 402). The practice of selecting general theories from other disciplinary fields and ‘domesticating’ them to fit the phenomenon under study have received some criticism (Davis & Marquis, 2005). This is because these imported theories operate at high levels of abstraction and are not able to effectively capture the variations in context of new and emerging forms of organising engendered by digital technologies. Rather than searching for a grand approach, Davis & Marquis (2005) suggest the development of theories that can help shed light on the causal paths that lead to observed phenomenon. However, the search for such causal processes remain theoretically grounded where general theory driven perspectives are complemented by the phenomenon specific causal explanations (Davis & Marquis, 2005; Salmon, 1998). The challenge of explanatory IS research involves constructing causal explanations with the context-dependent and dynamic interaction of people and technology. As a result, developing causal explanation requires cross-cutting multi-level analysis between the individual and the collective by establishing linkages between the same, making this an iterative and recursive process.

(c) Collecting and organising data from the case study as process research

The study involved overt non-participant observation of the team for a period of 11 months (15 July 2021 – 16 June 2022). Non-participant observations are extensively used in case study research and help gain an understanding of the phenomenon in its natural context (Mills et al., 2010). As a non-participant observer, the researcher did not participate in the activities of the group but the participants were aware of their presence in their midst for research purposes. The nature and purpose of the research was explained to the team at the beginning of the research process. In the line with evolving nature of the enquiry the observation involved a three-stage funnel process: descriptive observation stage to gain an overview of the setting, focused observation by narrowing observation down to areas of theoretical interest, and selected observation or observing relations among elements so identified i.e. the nature of digital architecture, resource portfolio, relationships within the proximate context, and wider environmental drivers (Spradley, 1980).

As a non-participant observer, the researcher shadowed the team in their team meetings, partner meetings, project meetings, including any seminars and workshops organised by them or to which they were invited. This also included presence in the team's Slack channel which was used for informal communication and sharing of quick updates. Documents collected over the course of this time included forecasting reports, project reports, academic publications, and software documentation including code repositories on GitHub. Given overlaps with the pandemic, presence involved both virtual and in-person. Projects and partner meetings were exclusively virtual given the involvement of multiple stakeholders from different countries and locations. These diverse data sources were first organised chronologically in phases. At this stage, for each phase, associated stakeholders, functional extensions, and choices were identified. Choices represented the objective outcome or definitive descriptor for each phase. These essentially represented the transmutation of ecosystem and environmental dynamics in relation to the technical architecture. This was presented to the team for verification and minor course correction. Sequence of events prior to the year of observation was according to the publication year of the paper or report – this provides a close approximation as academic papers in computer science tend to have a quick turnaround time from submission to publication. The second iteration involved identification of resources associated with each phase. This highlighted the need to pin down resource categories as well as engage in definitional work around what each resource category represented. Environmental conditions were elicited from observations followed by independent research into social conditions like the refugee crisis in 2016 or policy requirement like ethical use of automated systems, or flashpoint events like the war in Ukraine. These were confirmed through further observations and conversations with the Group and or their partners. This research was also the recipient of the LSE Knowledge Exchange and Impact fund which allowed the bringing together of multiple stakeholders under one roof which helped re-confirm the contextual and organisational considerations implicated within the model development process. The third iteration involved tabulating outcomes according to resources, relationships and structural conditions, for which categories were iteratively defined. These findings were organised in successive phases and tagged P1, P2, P3…. Pn (Appendix 1). These were then analysed using the theoretical framework to elicit portable concepts for theoretical contribution and generalisation within existing research.

Table 1. Data points (Corpus: Appendix 4)	Total
Team meetings	50
Meetings with external partners:	
Humanitarian organisations	6
Others e.g. academic collaborators	8
Project Meetings	
HiDALGO	25
VECMA	11
ITFLOWS	19
SEAVEA	8
Workshops	7

Internal documents	17
Publications and project deliverables:	
Journal articles, conference papers, dissertation, and PhD thesis	21
HiDALGO project reports	7
VECMA project reports	6
ITFLOWS project reports	15
Preliminary interviews	5
Total	**205**

(d) First person narrative account of field work and the challenges of doing process research

There were diverse data points that made up the Flee ABM case study. As mentioned earlier, data points included field notes, scientific publications, projects reports from institutionally funded projects and scientific consortia like HiDALGO, VECMA, ITFLOWS, SEAVEA, internal documents, and preliminary interviews. Entering the field was preceded by a preliminary conversation with the team lead. During this conversation I discussed the purpose and motivation for doing the research and the potential level of access I was hopeful for as well as the potentially year-long timeline of observation. At this meeting, I struggled to understand the technical description of Flee (a struggle that was going to continue for the initial stages of field work) but grasped that they had collaborations with humanitarian organisations like Save the Children. When he mentioned they were due to have a meeting in a few days, I requested to be present in that meeting to which he agreed and I received a calendar invite. At this initial meeting, I also highlighted the need to better understand the process and technology as a result of which he shared his student's PhD thesis as well as three journal articles to provide an overall understanding of the ABM computational technique and approach; they were Suleimenova, 2020; Suleimenova et al., 2017a; Suleimenova & Groen, 2020; and Suleimenova et al., 2021 (Appendix 3). The first meeting with Save the Children took place on 04/05/2021 and preceded the formal beginning of the field work on 15/07/2021. Along with the preliminary interviews that were conducted in the initial stages of the field research this meeting helped orient me to how the development of Flee was responsive to its context. This included the constraints posed by the technical architecture and the proposed approaches to overcome such constraints, thereby signalling inputs that were taken into account in the model development process. The initial discussion with Save the Children was also grounding as it was an instance of collaborating with a non-technical partner where issues under discussion were relatable to me having worked in the domain of ICT and socio-economic development in India. Issues like organisational technical capacity, sustainability and funding, usability, recruiting insights from dispersed field teams were familiar as organisational considerations for integrating technology within the process of long term socio-economic development.

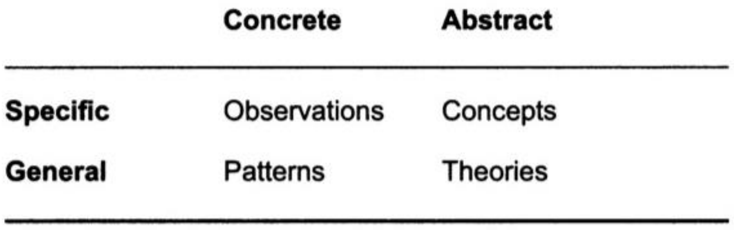
The field work formally started when I attended the first team meeting on 15/07/2021. This was in-person and had the full team in attendance. The team lead had already mentioned to the team about my participation. At this meeting I introduced myself and research motivation and timeline for research. Each team member introduced themselves in turn and the work that they do with respect to the development of Flee. The first team meeting and many successive meetings since took place out in the open in Crank Gardens at BUL due to COVID-19 considerations until eventually winter forced us online or inside. On this day of the first meeting, after the team meeting was over, I had the opportunity to talk to team members individually. This helped provide an insight into how the development of Flee was organised across different institutionally funded scientific consortia and how different work streams within particular projects were organised into 'work packages'. It also helped get an overview of the nature of collaboration with different humanitarian organisations over time as well get a sense of the overall development of the Flee model to date. This meeting also helped in asking questions around technical terms like sensitivity analysis that I had encountered in the articles and reports I had read which helped in understanding their relevance for the model. The initial meetings and technical workshops attended were challenging to the extent that in the first technical tutorial workshop I attended online for VECMA, I could not follow enough to take notes. Over time I was able to follow team meeting more clearly in terms of the substance and content of the meetings as my understanding of computational approaches and terminologies progressed. For example, after a few meetings of encountering the term 'ensemble models' I was able to ask clarifications of the same and how they related to some of the IS literature that referred to them (Leonardi et al., 2021). This also helped understand how they formed a part of a wider suite of approaches within simulation of forced displacement. Some of the more software oriented project meetings like VECMA, HiDALGO, and SEAVEA (VECMA's successor project) were quite technical. Over time I developed a strategy of understanding how particular specialisations or expertise contributed towards a given outcome which were reconfirmed through project reports and publications.

The frequency of team meetings varied; on an average they happened once a week but during busy periods they happened twice a week except during annual leaves. Busy periods were around start of the academic term or around deliverable deadlines. Sometimes meetings with humanitarian organisations threw up new requirements specifications which would also indicate an intensive phase of work to quickly develop working models. Team meetings were regularly attended by all team members except in cases of exigencies and typically lasted between an hour to hour and a half. Project meetings usually happened every two weeks and I attended only those ‘work package’ meetings within the projects of which the BUL team was a part. These work package meetings were for members of a given consortia involved in a given workstream but meetings usually only saw a selection of members attending. This depended on the updates to be shared and could vary from being as short as 10-15 minutes to sometimes over an hour. Most of the time project meetings involved sharing of progress updates except some like ITFLOWS which saw multi-disciplinary attendance from user boards, legal and programme teams, and developers to discuss how these multiple demands could be managed and met. When project developments were discussed in detail in a team meeting it reinforced their relative importance for the development process. For example, when ITFLOWS released a report on migrant journeys this was discussed in relation to the ongoing iteration of the Flee rule set of assumptions and parameters that governed the behaviour of the model. Field notes were organised with serial number and date in OneNote and sorted according to categories in which they belonged. Categories included Team meetings; Partner meetings with subcategories Humanitarian organisations and Others (external academic collaborations); Project meetings with subcategories HiDALGO, VECMA, ITFLOWS, and SEAVEA; and Workshops.

Iterative back and forth between journal articles, projects reports, and field notes helped establish at a first level a sense of evolution of the model. A full corpus of journal articles were then iteratively collected from the beginning of first version of the Flee model until the end of the observation period. These were then sorted date wise according to year of publication as were the projects reports. Projects reports were collected from the project websites, technical projects like HiDALGO and VECMA offered a straightforward identification of projects reports related to the work packages. However, given the interdisciplinary nature of ITFLOWS, data corpus included a wider set of documents to understand the relevant interdependencies and cross-fertilisation of ideas. At the end of the fieldwork period a simplified process timeline was developed which included the functional phases, associated choices and stakeholders involved. The timeline was constructed from 2016 i.e. the first version of Flee until the end of the fieldwork period i.e. 16/06/2022. The fieldwork stopped at a point of theoretical saturation where no new insight could be gleaned even through the model development process continues to proceed at the time of writing. The timeline was based on the publication dates of the documents as a proxy for when particular developments concretised over time. This simplified timeline was presented to the team to ensure correctness and receive feedback. Subsequently the organisation of data into process research involved iterative refinement of the form of presentation. In the very first iteration different inputs were identified; an overview of these inputs helped develop resource categories. These resource categories were then mapped onto each phase in the evolving iterations of organising the data. While the participants in each phase were identified in the initial iteration, reengaging with the data helped relate relationships to resource categories. Relevant environmental conditions were identified through reports and observations; this was followed by independent research and constantly relating them with the data corpus which helped in identifications of key environmental drivers which were then mapped onto each phase, providing an overview of the evolution of the Flee ABM and helped elicit a description of the nature of the digital architecture and interdependencies with its social components of the ecosystem and the environment. The nature of architecture helped understand the inter-relationships between actor, structure, technology and task within the proximate and distant context. The next step was to compose a narrative based on this data organisation and insights. Working to develop the narrative foregrounded underlying conditions like performance testing and evaluation which though did not culminate in key phases nevertheless improved the model. The narrative work itself helped clarify interlinkages and insights between how the analytical constructs within the adapted choice framework translated to explanations for relationships between constructs of theoretical interest. The iterative process of constructing and reconstructing the narrative helped link the surface and deep structure and derive explanations from descriptions.

4.2 Of what is this a case? Moving from observation to theorisation

Case study research is an evolving enquiry and it is only when the researcher is quite deep into the research process that they are able to understand the kind of work a case is likely to do. Thus, the *a priori* constructs form a first step in a “long series of gradual precisions” (Lund, 2014; p. 227). Lund (2014) argues that case studies are often presented as self-evident and what the material is a case of is less evident. A case is “edited chunk of empirical reality” where particular features are ‘marked out, emphasized, and privileged while others recede into the background' (p. 224). Therefore, a case becomes a construct for "organizing knowledge about reality in a manageable way" (p. 224). Analytical movements of generalisation, specification, concretisation, and abstraction help to define the case and potential area of contribution more clearly. To know what case studies are about it is useful to locate their contents across the continuum from specific to general and concrete to abstract.



	Concrete	**Abstract**
Specific	Observations	Concepts
General	Patterns	Theories

Fig. 2: Lund's (2014) Analytical matrix (p. 225)

The continua between specific and abstract are open ended rather than two binary opposites as shown in Figure 2. Configurations of specification, generalisation, concretisation, and abstraction define the boundaries of the case. Specification involves identifying an event or set of events for study and cases are concrete in the sense that they actually happened. This involves the development of thick causal concepts that help to move from data to empirical statements (Cartwright, 2004; Lee & Baskerville, 2003). However, to move from empirical statements to theoretical statements and contribute to theory the study needs take up considerations for generalising, abstracting, and theorising. Once thick causal concepts have been defined on the basis of specific and concrete events, the movement from empirical to theoretical statements are guided by mediating conditions. Moving towards theoretical statements involve highlighting the general significance of such events. Generalisation can be empirical and analytical: empirical generalisation is an extrapolation or claim that the knowledge gleaned from a limited number of events holds true for a larger group; analytical generalisation on the other hand refers to the identification of constituent elements or properties of an event. Data from a multiplicity of sources provide the basis for observation of patterns and mediating conditions within the corpus. Generalisation beyond the context of the study does not mean to overreach and universalise with the purpose of making universal declaration alongside caveats to cover oneself. Generalisation means entering into a dialogue where one's research resonates with other works. These are not to establish actual validity but a likelihood of providing provisional propositions.

Lee & Baskeville (2003) offer a framework of 4 different types of generalisability i.e. generalising from data to description, from empirical observations to theoretical constructs, from theoretical constructs to lessons for practice, and 'pure theory' generalisations from theoretical constructs to theoretical constructs. It can be argued that these four approaches to generalisation form a continuum from empirical observation to abstraction and theorisation. Thereby generalisation becomes an iterative process of developing a knowledge claim which is already true for a particular setting and can be potentially be true in other clearly defined settings (Seddon & Scheepers, 2012; 2015). This involves understanding the distinction between empirical and theoretical statements and having clarity about generalisation from and generalisation to (Yin, 2014). Therefore, theoretical contributions involve a series of abstraction guided by parsimony to identify qualifying conditions for theory development (Eisenhardt & Graebner, 2007). Within this study, these qualifying conditions for theory development corresponds to the building blocks (thick descriptions and mediating conditions) that enable analysis and development of theory.

While the point of departure may have been specific and concrete, any study is analysed through a set of concepts which help in abstracting and editing the data by making its particular inherent qualities prominent (Lund, 2014). Choice of concepts partly define a case. A research study does not discover new events but novel ways of thinking about connections and relations that are not directly observable about a known phenomenon (Danermark et al., 2002). Abstraction is process of creative reasoning to formulate new ideas about a particular phenomenon, to see relations and connections that were not obvious and to think about something in a different context (Danermark et al., 2002). Preferred *a priori* concepts are always abstract and research is obliged to track and explain the movements between concrete manifestations of abstract phenomenon and be explicit about how concepts are operationalised. While theoretical questions help deduce critical areas of enquiry, field research helps to induct empirical observations to explore concrete dynamics. Going back and forth between observations and abstract concepts helps in progressively approximating both by helping to better discern the phenomenon empirically and describe it conceptually.

Generalisation is an attempt to see resonance with events and processes in the same level of abstraction but in different contexts while abstractions aim to identify the decontextualised qualities or properties in studied events. Theorisation then becomes the process of moving from empirical observations through concepts to be able to say something about inherent qualities and dynamics of events in contexts other than the one being studied. This involves both a decontextualised abstraction and transfactual corroboration (Lund, 2014). However, methods employed to traverse this continuum towards generalisation are not as systematic as comes across in research studies and orderliness is always established in hindsight (Feyeraband, 1975). The research process itself involves the constant and sometimes aimless back and forth between observations, generalisations, abstractions, and theorisations rather than a neat trajectory. However, it is only through this movement to and from theoretical questions and detailed observations that one is able to define the problem and explain it. It is this constant movement and their articulation that helps make analytical and theoretical contributions.

5. CASE DESCRIPTION AND FINDINGS

Chapter summary: *This chapter begins by providing an overview of Flee's development process. It then goes on to examine in detail the key elements of its architecture, ecosystem, and environment. It traces its journey from 2016, in the backdrop of the European refugee crisis, as a simple hard coded model of a single historical conflict created by a lone developer to its evolution as a computational model for a range of conflicts subtended by sophisticated computational and mathematical frameworks, a specialised team, funded scientific consortia, and sectoral collaborations that drove the innovation process. Unpacking this process provided insights on the nature of the technical architecture that needed to be negotiated, specialised multi-actor configurations that had to be formed within the ecosystem, and the attendant environment drivers that are implicated within the process. It identifies and defines the resource categories (computational, data, knowledge, information, financial and human) used and developed during the course of this journey, the hybrid nature of the architecture with its integrality supported by generic software components, the nature of relationships and their configurations within the ecosystem, and the diverse environmental drivers (computational, digital and data; domain; socio-political; and policy, legal and ethical) that stand in circular inter-locking relationship with each other.*

5.1 The case narrative: An overview of Flee ABM development

Flee is an ABM modelling toolkit written in Python and purpose built for simulating the movement of individuals across geographical locations. It is available open source (https://github.com/djgroen/flee) under a BSD 3-clause license (https://github.com/djgroen/flee/blob/master/LICENSE). Like ABM models Flee is primarily a computational model that simulates the actions of autonomous agents. This helps arrive at the overall behaviour of a complex system based on rule sets comprising of parameters and assumptions which defines how agents act or interact with their environment. Flee was built for simulating forced displacement where each agent within the model represents a forcibly displaced person whose decision to 'flee' across geographical locations, i.e. from a location of conflict (conflict zones) to a location of safety (camps), is governed by a set of parameters and assumptions which make up its rule set. The purpose of this simulation was to predict and forecast the arrivals of forcibly displaced individuals across camps in neighbouring countries to aid humanitarian planning and operations.

Parameters and assumptions within Flee involve the likelihood of movement given a particular location (depending on whether it is a conflict zone, transit point, or camp), maximum movement speed (depending on the mode of transport), and mode of transport (walking or shared vehicles) (Groen, 2016; Suleimenova et al., 2017a; Suleimenova et al., 2021; Internal documents; Field notes). While these straightforward parameters and assumptions within the rule set determine the movement of an agent from one point to another, given the multicausal nature of forced displacement a number of conditionals have also been introduced over time. These include camp and border closures leading to forced redirection (Suleimenova et al., 2017a; Suleimenova, 2020; Suleimenova & Groen, 2020), food security (Campos et al., 2019), weather conditions (Jahani et al., 2021), mining and mineral pricing (Groen et al., 2019a), and route accessibility (Boesjes et al., 2022; Boesjes, 2022; Field notes) – all of which determine the movement of forcibly displaced individuals and the distribution of their arrivals across camps over time. While in the initial stages, parameters and assumptions in the rule set were based on intuition (Suleimenova et al., 2017a), they have evolved over the years through qualitative feedback and information received from humanitarian organisations on how conditions unfold on the ground (HiDALGO, 2021b; Internal documents; Field notes).

The development of Flee is anchored by the team at the Department of Computer Science at BUL. The team comprised of Derek Groen (DG), Diana Suleimenova (DS), Alireza Jahani (AJ), and Yani Xue (YX). The current composition of the team developed over a period of time. It began with DG who joined BUL as Lecturer from the University College London's (UCL's) Centre for Computational Science where he was a Postdoctoral researcher after having completed his PhD from the University of Leiden. DG's professional networks across the UK and Netherlands have led to introductions to and collaboration with funded scientific consortia (Interview with DG, 18/10/2021) that helped in sustaining the development of Flee through financial support, specialised technical expertise, and computational resources like supercomputers.

> *What happened in 2016, in the spring, one of the professors from my old university said "Well, Derek you should get in touch with CoeGSS Centre of Excellence," which is basically geospatial systems centre of excellence, they do some agent based modelling so there was some opportunity for collaboration there. So I contacted them.*

> *They were very interested but couldn't do much at the time because that project had its own trajectory. What happened half a year later is that they had a different proposal for HiDALGO (Interview with DG, 18/10/2021).*

DS joined as DG's PhD student in 2016 which significantly advanced efforts towards the development of Flee. The two foundational aspects of Flee in the form of a generalised Simulation Development Approach (SDA) and automation of workflows, that would become the cornerstone of Flee's further development, were a part of her doctoral thesis. After the conclusion of her PhD in 2020 she remained a part of the team as a Postdoctoral researcher and eventually a Lecturer in the same Department primarily involved in sensitivity analysis and exploratory work on IDPs, forecasting conflicts evolving in real time, and providing support on rule set updates. There was another Postdoctoral researcher, Hamid Arabnejad (HA), who was a part of the team for 3 years before moving on in 2021. HA's expertise was in High Performance Computing (HPC) and distributed and parallel computing. HA introduced the process of execution of Flee on supercomputers and provided critical software support and development. AJ joined in 2020 and worked on constructing new conflict scenarios and couplings with conditionals like weather and telecommunications data. YX joined in 2021 and brought her expertise in multi-objective optimisation (MOO) and evolutionary algorithms within the domain of machine learning and artificial intelligence.

The initial model of Flee was developed in 2016 by DG in the backdrop of 2015-16 migrant crisis (Workshop presentations by DG). Many aspects of its progress have been consolidated over time with the joining of new team members who took up different aspects of its development and formed the core unit responsible for coding, rule set updates, software support, construction of conflict scenarios, and integration and development of new computational frameworks and approaches (Field notes). Initially, the first version of the model based on the 2012 Mali conflict was hard coded including location names which made it very difficult to construct new scenarios or replicate the model across different contexts. In developing the first version of the model DG opted for simplicity over detail to establish a prototype. Going forward, to enable replicability across different contexts other conflicts scenarios were selected. These included Burundi, Central African Republic, and South Sudan with the objective of applying Flee in a way that could be easily adapted to different conflict contexts (Interview with DS, 15/07/2021; Suleimenova, 2020). This led to the formulation of a generalised SDA and automation of workflows, the latter was implemented with the help of an automation toolkit and streamlined architecture that accepted datasets in a pre-processed format. This enabled any changes with regard to different scenarios to be implemented through a command in the automation toolkit rather than through changes in the main code itself.

Flee was made generalisable across conflict contexts using the SDA (Suleimeonva et al., 2017a) and aspects of model construction were automated through an automation toolkit called FabSim3 (FabSim version 3), which was adapted for migration simulation in the form of a plugin called FabFlee (Suleimenova et al., 2017b). Both SDA and FabFlee were aimed at enabling clear and easily implementable steps for quicker simulation construction which would be required in humanitarian response. SDA proposed a six step approach from problem selection to analysis that allowed for model refinement by incorporating conditionals. SDA was proposed as a generalised set of steps that could be followed for modelling any new conflict and ensured a unified approach across modelling instances. FabFlee complemented SDA by providing multifunctional support in the form of automation of workflows and coupling with conditionals while also forming part of a suite of software tools that helped run the Flee code in parallel on supercomputers to test and maximise its performance (Campos et al., 2019; HiDALGO, 2019c; Groen et al., 2019a; VECMA, 2019; Suleimenova, 2020; Suleimenova & Groen, 2020; VECMA, 2021a; VECMA, 2021b; HiDALGO, 2021c; Suleimenova et al., 2021; Groen et al., 2021). However, despite automation of certain workflows, the construction of a new model was a research-intensive and manual process that involved in-depth understanding of a given conflict by trawling through research reports, news articles, and available data. This was because each iteration and experimentation in Flee have been within the context of particular conflicts each of which have their own dynamics (Field notes).

As the model progressed it aimed to incorporate additional factors and levels of detail that would bring the model in closer approximation with reality. This was introduced through ensemble modelling (Suleimenova & Groen, 2020) and three instances of coupling (Campos et al., 2019; Groen et al., 2019a; Jahani et al., 2021). Ensemble modelling was used to evaluate the effect of policy decisions like border and camp closures and involved running the initial model multiple times; each time with slightly different conditional parameters to understand the impact of particular assumptions on the outcome of the model. The first two instances of coupling in the form of food security and conflict evolution were developed by coupling together corresponding sub-models for food security and conflict evolution respectively that ran in tandem with the main Flee code to simulate the effect of these conditionals on the movement of individuals to safety. The third instance involved a multiscale model i.e. coupling two models of differing temporal and spatial scales to simulate their combined effect on forced displacement. This involved coupling a macroscale model that included most of the conflict country with a microscale model of a particular region within the country which in turn was coupled with weather data like precipitation and river discharge which affected the accessibility of the routes in the region. Since detailed simulation models for forced migration do not exist, these coupling scenarios served as initial prototypes to explore these relationships (HiDALGO, 2019b).

However, introducing these conditions and exploring their effects required technical development of multiscale models. This involved combining or coupling of multiple models covering different spatio-temporal scales and exploring their performance with the introduction of multiple details and iterations or expanding the scale of the simulation itself by increasing the number of agents. Multiscale modelling and coupling were supported through the development of the multiscale modelling approach as well as investigations into the scalability and performance of the code both on local desktops as well as remote supercomputers (Groen, 2018; HiDALGO, 2019a; HiDALGO, 2019b; HiDALGO, 2021a). Remote execution on supercomputers became particularly important when testing the efficiency of code, particularly when the number or level of detail were increased in the model. Further, the Flee output is highly sensitive to input parameters and assumptions. Therefore, in order to increase accuracy of the model it became important to understand which of the given parameters had an outsized impact on the output. This led to the Sensitivity Driven Simulation Development (SDSD) approach (Suleimenova et al., 2021). SDSD also presented the concretisation of the ongoing work on Flee 2.0 rule set (Internal documents; Feedback from first review of process timeline). The insights derived from implementing SDSD approach would help in examining the underlying logics within the model and make necessary changes to ensure the model was predicting to high levels of accuracy. The notion of accuracy mentioned within this study refers to optimisation of error rates. Error rates involve statistical and mathematical measures of deviation of the model output from datasets like the UNHCR. Approaches to measure error rates by validating the model output to UNHCR data involved a range of mathematical approaches from average relative difference to more formalised approaches for validation, verification, and uncertainty quantification (VVUQ) and sensitivity analysis.

However, mathematical parameter exploration was not the only way in which fundamental changes were made to the rule sets. As the model progressed, mathematical formalisation and computational rendition were the culmination of ongoing discussions, collaborations, and feedback from humanitarian organisations (Field notes). This was in contrast to numerical optimisation exercises which are commonly applied in the field (DG second review feedback). Humanitarian organisations provided the team with qualitative input that eventually translated into Flee 2.0 and Flee 3.0 rule set updates which provide the back-bone for the Flee code (Flee 3.0 was a work-in-progress when fieldwork concluded) (Internal documents; Field notes; HiDALGO, 2021b). The team has had varying degrees of collaboration with relevant departments in humanitarian organisations like UNHCR (United Nations High Commissioner for Refugees), International Organisation for Migration (IOM), and Save the Children.[1] However, collaborations with humanitarian organisations tended to be ad hoc with periods of intense collaboration. The team worked in tandem with UNHCR when they were developing their own machine learning model for forced displacement. UNHCR also provided a letter of support when the team was making the funding application for HiDALGO (Interview with DG, 18/10/2021).

> *It [HiDALGO] was a huge project and when I wrote that proposal I did not have any funding and I was thinking how can I get support. Then I met UNHCR … at some event. We had a few calls, I think that was probably before HiDALGO started and they provided a letter of support. That was the first time we had a connect with any NGO (Interview with DG, 18/10/2021).*

IOM facilitated a qualitative survey with NGOs on the ground to explore the validity of parameters and assumptions leading to the evolution of Flee 2.0 rule set (Internal documents; HiDALGO, 2021b). The collaboration with Save the Children involved forecasting the then unfolding Tigray conflict which was also the first time two significant areas of advancement was undertaken with respect to the Flee model – integrated modelling of refugees and IDPs

---

[1] The thesis henceforth refers to these collaborations with relevant departments by their organisational names or acronyms.

and forecasting (Suleimenova et al., 2022). IDP modelling was previously not undertaken due to lack of available data and datasets against which to validate model outcomes (Groen, 2016). However, in this case because of the collaboration it became possible due to the potential availability of greater insight (Field notes). Previously, models had predicted historical conflict to be able to test model performance through validation with existing UNHCR data, however, in an ongoing conflict it became important to forecast or predict forward based on conflict evolution scenarios[2]. This collaboration with Save the Children on forecasting and modelling IDPs also attracted IOM's attention during a workshop (Flee workshop held on 24/09/2021; Field notes) who then resuscitated the connection to explore the possibility of developing scenarios to overcome the non-availability of data as a result of lack of access due to violent conflict in the region (Field notes).

> In the meeting with IOM on 07/10/2021, they mentioned attending the Flee workshop on 24/09/2021. On knowing about the partnership with Save the Children, they did not want to be 'left behind' given the sectoral push to anticipatory action. They wanted to ensure that they were up to speed in forecasting and anticipatory management (Field notes - Humanitarian Organisations #3).

> The follow-up meeting with IOM on 12/11/2021 with country and data specialists highlighted how tough conflict and displacement predictions could be. If one of the factions within a given conflict were to get an upper hand, that could completely reshape conflict dynamics. The situations are very volatile and unpredictable. Depending on how a conflict evolves, this could potentially result in different scenarios - either concentration of troops in border areas, or march towards the capital, or talks and negotiations. All of which have varying degrees of impact on the extent of forced displacement. Due to the rapidly changing conflict dynamics, by the time data can be accessed, cleaned, and put down for analysis it becomes quickly outdated. Given these data gaps, possibilities were explored of sandboxing with certain scenarios which can give the analyst a few options for planning and management. For e.g. if data cannot be accessed from anywhere, be it remote sensing or density counts, then simulation scenarios can serve as a substitute for a piece of the puzzle (Field notes - Humanitarian organisations #4).

---

[2] The terminology of prediction and forecasting was used by the team to distinguish between 'predicting' movement in historical conflicts and 'forecasting' future or evolving ones.

While inputs from humanitarian organisations enriched the model development process, the ongoing development of Flee would not have been possible without the funded scientific consortia of which it has been a part. Though the initial development of Flee was made possible through DG's existing professional networks, the joining of DS as a PhD student, and a carry-over grant from his time at the UCL (Computing Patterns for High Performance Multiscale Computing – ComPat), substantial financial investment was required to expand both the team and technical capacities. These funded consortia were instrumental in sustaining the development process not only through financial support but also with resources like supercomputers and a wider pool of expertise that facilitated cross-collaboration. HPC and Big Data Technologies for Global Challenges (HiDALGO, https://hidalgo-project.eu/) project aimed at developing novel approaches in HPC and High Performance Data Analytics (HPDA) to respond to critical global challenges in which Flee was the use case application as the migration pilot. DG's former colleague from the University of Amsterdam was instrumental in making the introduction to Centre of Excellence for Global Systems Science (CoeGSS), considered a precursor to HiDALGO, at the proposal stage which helped him get on board with this consortium (Interview with DG, 18/10/2021).

Verified Exascale Computing for Multiscale Applications (VECMA, https://www.vecma.eu/) was aimed at enhancing the reliability of computer simulations in critical application sectors like migration by developing a suite of software tools for VVUQ. Both VECMA and HiDALGO ended in 2021 / 2022 and had received funding from the European Union's Horizon 2020 framework. They were long term multi-year projects supporting a range of development activities highlighted above. While collaboration with the expertise within a consortia led to some of the key developments like multiscale simulations and SDSD, it also sustained longer term development processes like developing a parallelised version of the Flee code so that it could run simultaneously on the supercomputer and local desktop to increase efficiency, developing a generalised HPC workflow, and benchmarking and scalability on supercomputers to enable efficient execution of the code – all which sustained ongoing code development. VECMA's successor grant SEAVEA (Software Environment for Actionable and VVUQ-Evaluated Exascale Applications, https://www.seavea-project.org/) helped continue the progress on VVUQ suite of application tools and also included Save the Children as one of the consortium partners. Each of these projects had common partners between them with some associations that have continued for years across successive grants where various collective scientific advances were made under different funding mandates.

IT Tools and Methods for Managing Migration Flows (ITFLOWS, https://www.itflows.eu/), also funded under the EU Horizon 2020 Framework, was a multi-disciplinary project at the intersection of migration, technology, and human rights. It aimed to forecast migration arrivals within the European Union to assist in humanitarian planning through the EUMigraTool (EMT) which comprised of two complementary approaches: the small scale model and the large scale model. The team was responsible for the small scale simulation model, Flee, which forecasts flow of forcibly displaced persons to neighbouring conflict countries. Another technical consortium partner was responsible for developing the large scale machine learning model that uses the output of Flee as an input and combined it with a host of other workstreams on social media and sentiment analysis to produce global forecasts for the EU. Within ITFLOWS, Flee aimed to explore short range migration patterns of forcibly displaced persons across a wide range of new conflicts, incorporating important demographic aspects such as gender, age and ethnicity (Field notes). The conflict scenarios within the project were to include Nigeria, Venezuela, Afghanistan, Syria, and Mali with Afghanistan being replaced by Ukraine as the outbreak of war dominated headlines. This was compounded by the difficulty in constructing the Afghanistan scenario due to lack of data as a result of limited presence of humanitarian organisations (Field Notes). Due to its multi-disciplinary approach and composition of the consortia, the ITFLOWS project foregrounded the tensions between legal and ethical expectations and technical constraints to ensure appropriate terminologies for predicted or forecasted outputs with regard to legal validity and to enable actionable response (Field notes).

> The first ITFLOWS meeting stressed on the multi-user environment of its predictive tool and discussed how to deal with errors within that environment. This highlighted the need to critically engage with the quality and feasibility of the data and to engage with uncertainties that come with data collected for different purposes. This also raised questions around where such a tool would be hosted (Field notes - ITFLOWS #1) with caveats from prospective users ranging from humanitarian organisations to local administrations that it should not become the basis for restrictive policies and that it should be used for dignified reception and access to resources (Field notes - ITFLOWS #2).

> The subsequent meeting highlighted how different datasets, originating from different jurisdictions, institutions or organisations, used in the modelling process contained different definitions of the term ‘refugee’. Since the modelling process used a mix of different datasets, it became difficult to determine how the notion of the term ‘refugees’ transmuted over the course of the modelling process into the modelling output. This can be encapsulated in the following paraphrased statement: *if any expert in international law asks who is a refugee in your model, the entire model collapses* (Field notes - ITFLOWS #3).

This speaks to the wider environment within which the development of Flee has unfolded. It was first developed within the context of the 2015-16 migrant crisis. Throughout the course of its evolution, the demands for predictive analytics gathered momentum with ‘anticipatory action’ becoming the order of the day to maximise limited operational resources by prepositioning them where they were needed the most (UNDRR, 2015; Lowcock, 2019). Anticipatory action refers to action taken in advance or anticipation of an impending disaster or crises to reduce its impact when it occurs. Forecasts or predictive analyses trigger the actions (usually when a given threshold is breached), in the form of prearranged financing and implementation steps (IFRC, 2020). This was unfolding in the wake of a move towards UN 2.0 Quintet of Change driven by a 5-point agenda leveraging data and digital transformation for system-wide changes (UN, 2021). However, in parallel, there were also concerns about the use of predictive analytics as a result of their potential for bias, discrimination, and misuse which could push already vulnerable people into precarity (UNOCHA, 2021), this also includes biases and gaps in the data that is available to work with (UNOCHA, 2018). Further, as the ITFLOWS project also showed, that alongside ensuring fair and ethical processes there was also a need for harmonisation with a range of legal systems and frameworks when it comes to developing actionable insights for forcibly displaced persons to ensure that there is legal validity of and congruency between the legal terminology and computational output given the different approaches taken by different agencies and administrations towards the reception, processing, and rehabilitation of displaced persons (ITFLOWS, 2021a; 2021b; 2021c). Environmental conditions vary from such longer term policy expectations, to flashpoint events like Tigray, Ukraine, and Afghanistan (Redfern, 2022), to more pervasive issues around missing data and the refugee registration numbers published by UNHCR (KEI Event, 2022). The latter keeps changing in the aftermath of review exercises resulting in a revision of previously published numbers; thereby complicating validation efforts that enable the calculation of error rates which determine the accuracy of the model (Groen, 2016; Chan et al., 2018). Moreover, each conflict has different dynamics and sometimes despite knowing particular motivating factors, the relationship cannot be simulated due to lack of adequate data required to define and validate the relationship (Suleimenova et al., 2017a; Field notes). However, the conditions are not just limited to the social, political, and policy environment but also the funding mandates within which it has to operate which highlights a pervasive tension within the development of Flee between competing objectives and actions and the trade-offs that have to be resolved in order to progress towards each successive stage.

5.2 Architecture: Key phases in the evolution of Flee

The following sections highlight key phases in the evolution of the Flee architecture. While the Appendices 1 and 2 tabulate key phases as discrete happenings at a point in time, reality is more complex than a linear chains of events. It is a circular, mutually reinforcing, and mutually informing process. Therefore, the process timeline in Appendices 1 and 2 are an attempt to simplify dynamic reality that help unpack constituent elements that result in a particular outcome within given phases while taking collective outcomes as a point of departure for analysis. The development of the initial model (P1), generalisation of a replicable SDA (P2), and automation of certain workflows (P3, P9, P12) provided the foundational basis for further development of Flee and remained a critical component that helped anchor newer direction of development. Subsequent developments like the parallelisation of the code (P5, P11) and development of the coupling approach (P7, P8, P10) helped in scaling the code to supercomputers, helping to test its efficiency and performance, and exploring the role of additional causal factors like food insecurity, conflict evolution, and weather conditions. Over time, sensitivity analysis came to be an integrated aspect of the SDA through the SDSD approach which reflected formalisation of the Flee ruleset update (Flee 2.0) (P13). This formalisation also helped incorporate feedback from humanitarian organisations. Ongoing collaboration with a humanitarian organisation translated into the first attempt at forecasting a conflict unfolding in real time while also incorporating IDPs within the modelling process (P15) which eventually facilitated the evolution of Flee 3.0. There were also attempts to expand the repertoire of problems addressed by Flee. This included resolving the challenge of choosing an optimal camp placement location by deploying the simulation-optimisation approach combining Flee in conjunction with an evolutionary algorithm like NSGA-II (P14). These developments over time were anchored through funded scientific consortia like HiDALGO, VECMA, ITFLOWS, and SEAVEA (P7 onwards). However, there were also scientific enquiries, which while beneficial, failed to find traction like the development of a gaming user interface for Flee (Estrada et al., 2017) (P4) due to lack of further incentive for collaboration (First review feedback). While the approach to test the feasibility of input data was an important one for understanding the limitations of such data (Chan et al., 2018) (P6), its core approach for predicting displacement counts was not incorporated as a standard practice due to systemic issues with datasets like UNHCR. However, as of 2022 there are now methods being tested to support the prediction of displaced persons, given a conflict evolution (DG second review feedback). The following sections group the different phases identified in the process timeline (Appendix 1) under key functional or developmental categories rather than in a linear narrative to provide a clearer overview of the key aspects of Flee's development.

5.2.1 Initial model

The first version of Flee (P1), developed in 2016 by DG, was based on the 2012 Northern Mali conflict (Groen, 2016). The aim of this first version was to explore the pattern of refugee movements or the distribution of refugees across camps in neighbouring countries as they flee conflict locations. The objective behind developing this model was to enable humanitarian and other support organisations to better prepare for their arrival, help governments understand the implications of deploying different border and immigration policies, and supplement the data deficits around the phenomenon of forced displacement. Particularly because simulation modelling can be helpful in refugee settings when available data on refugees makes it difficult to formulate causal inferences.

The model used camp location and refugee registration data from UNHCR (data.unhcr.org) and used Bing Maps (maps.bing.com) to determine distances for the path traversed by refugees in their journey to the camps. Refugees were spawned in different conflict locations and locations were interconnected with paths derived from Bing Maps. Refugee agents moved with a probability of 1 (certainty) from conflict locations, with the average refugee staying in a camp for 1000 days. The numbers used to populate refugee agents in source locations were drawn from UNHCR refugee registration numbers from destination camps. A moving refugee agent chose their destination based on a weighted probability function of 1 divided by route length. Travel distances between locations were estimated from Bing Maps as the shortest route required by cars with the assumption that refugees stick to major roads and use shared vehicles. Another assumption involved refugee movements taking one day as travel times for refugees were not available. The data was analysed using a Pandas library (pandas.pydata.org) and visualised using matpotlib.

To validate the model, the simulation results were compared to UNHCR data and as well as by calculating the sum of absolute differences in refugee counts as a proportion of the total number of refugees. The results of the simulation could not be reconciled with the UNHCR data due to changes in the data after recounting following which refugee counts with UNHCR decreased by 68%. There were also inconsistencies in arrival numbers between simulation result and UNHCR data that could be explained by differences in registration approaches between different camps in neighbouring countries which could in turn be the reason for under-reported refugee arrivals in a particular location. Such revisions in UNHCR datasets continued to complicate validation exercises and optimisation of error rates throughout the model development process.

Some of the challenges involved the complex nature of civil wars and the lack of systematic reports which made it difficult to develop accurate models of refugee movements. However, this first version demonstrated a proof-of-concept for developing simulation models for forced displacement as well as pinpoint the specific challenges and opportunities. This relatively simplistic version of the model helped to form the foundational basis for more advanced explorations in the area of refugee modelling using HPC that offer high degrees of scalability and the ability to use multiple approaches within ABM simultaneously.

5.2.2 Generalisation

The development of Flee, which was hardcoded in the initial phase, was generalised through the SDA for three conflicts in Africa – Mali, Burundi, and Central African Republic (Suleimenova et al., 2017a) (P2). One of the main reasons for developing the SDA was the need for humanitarian organisations to facilitate rapid simulation development when a conflict strikes. SDA suggests six phases for simulation construction as illustrated in Figure 3. They are based on problem formulation (Phase 1), translation to a computational model through data source selection, model construction and model refinement (Phases 2, 3, and 4), and operational validation through simulation execution and analysis (Phases 5 and 6). SDA was able to help translate the different processes, like selection of data sources, extraction and conversion of data, and validation, into a step-wise approach that could be replicated across different conflict contexts to develop relevant simulation models. The approach involved processing input and validation data for refugees, constructing network graphs, and choosing simulation parameters and assumptions. In addition to UNHCR and geospatial data, SDA also incorporated the use of ACLED (Armed Conflict Location and Event Data Project) data to determine conflict locations.

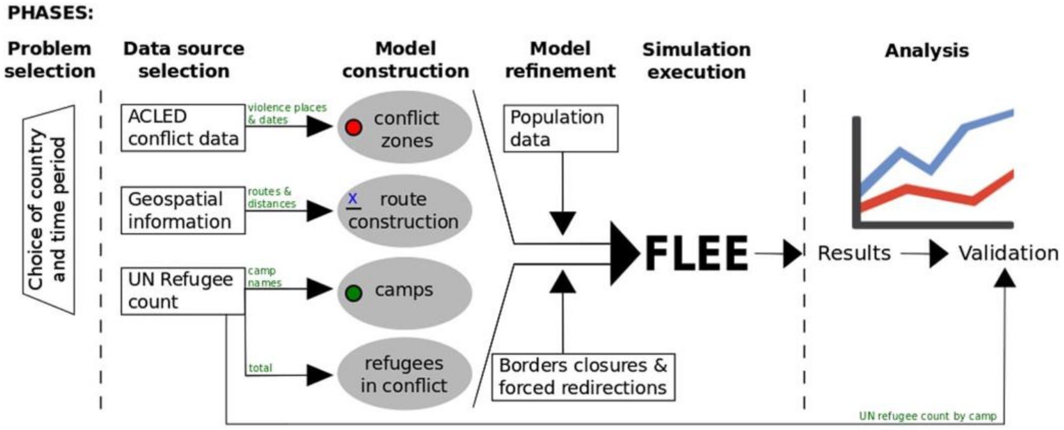


Fig. 3: SDA for predicting distribution of refugee arrivals across camps. (Source: Suleimenova et al., 2017a)

Each time-step within the simulation represents a day and at each step refugees are inserted into the simulation based on the daily increase in total refugee registration count from the UNHCR data. These refugees are then inserted into their location of origin which relates to the conflict location obtained from the ACLED database. The exact location is picked from among all conflict zones with the probability of a location being selected proportional to its population. The population of a location decreases each time a refugee agent is created. Since refugee agents are spawned in conflict locations on the same day as camp registrations and travel is non-instantaneous, the output generally results in under-prediction of refugees which is corrected by multiplying the refugee population in a given camp with the ratio of the total number of refugee count for the conflict on a given day according to UNHCR and the simulation output of total number of refugees in camps on the same day. Decreases in UNHCR refugee registrations create a ‘refugee debt’ variable which needs to be compensated by subsequent increases in registration numbers before additional agents are introduced in the conflict, particularly because agents are not deleted.

During each time step the refugee can travel zero, one, or more links which is determined by the probability of a refugee agent moving from a given location which is pegged at 1.0 for refugees in between transit location, 1.0 for refugees in conflict locations, 0.001 in camps, and 0.3 for all other locations. The assumptions at this stage were based on intuition due to the absence of empirical data or information. Parameters were not optimised to avoid overfitting which could reduce applicability to other contexts. When an agent traverses a link it needs to choose one of the available travel paths, path selection is done by a weighted probability function where weight is determined by attractiveness value divided by the length of the link in kilometres. The attractiveness value is 0.25 for conflict zones, 1.0 for other locations in the country, and 2.0 for locations abroad. Again, these values were based on intuition and the sensitivity of such assumptions were tested using sensitivity analyses to understand the contribution of particular parameter selection to error rates. Figure 5 illustrates the algorithmic assumptions and flows used in this phase of the Flee model. However, sensitivity analysis was not an integrated aspect of the SDA at this stage. It was found that none of these parameters had a significant effect on accuracy of simulations, neither did the assumption that refugees do not travel more than 200 kms per day. It was found that the simulation had low sensitivity to higher travel distances and the error increased if lower travel limits were adopted. It was assumed that the refugees took the shortest routes while travelling as determined by route planners like Bing. Figure 4 summarises how the generalisation phase improved upon the initial model through a comparison of the Mali model.

Aspect of simulation construction	Now	2016 ICCS paper
Number of conflict locations in simulation	11	3
Source of conflict information	ACLED	Wikipedia
Number of locations in Mali	16	9
Number of locations (total)	24	17
Correct data for mismatch between level 1 and level 2 registrations	Yes	No
Awareness range	1 link away	path distance only
FLEE version	1.0	0.1b (non-public beta)

Fig. 4: Summary of differences with respect to the Mali simulation in comparison with the initial model: Groen, 2016. (Source: Suleimenova et al., 2017a, Supplementary Note 5)

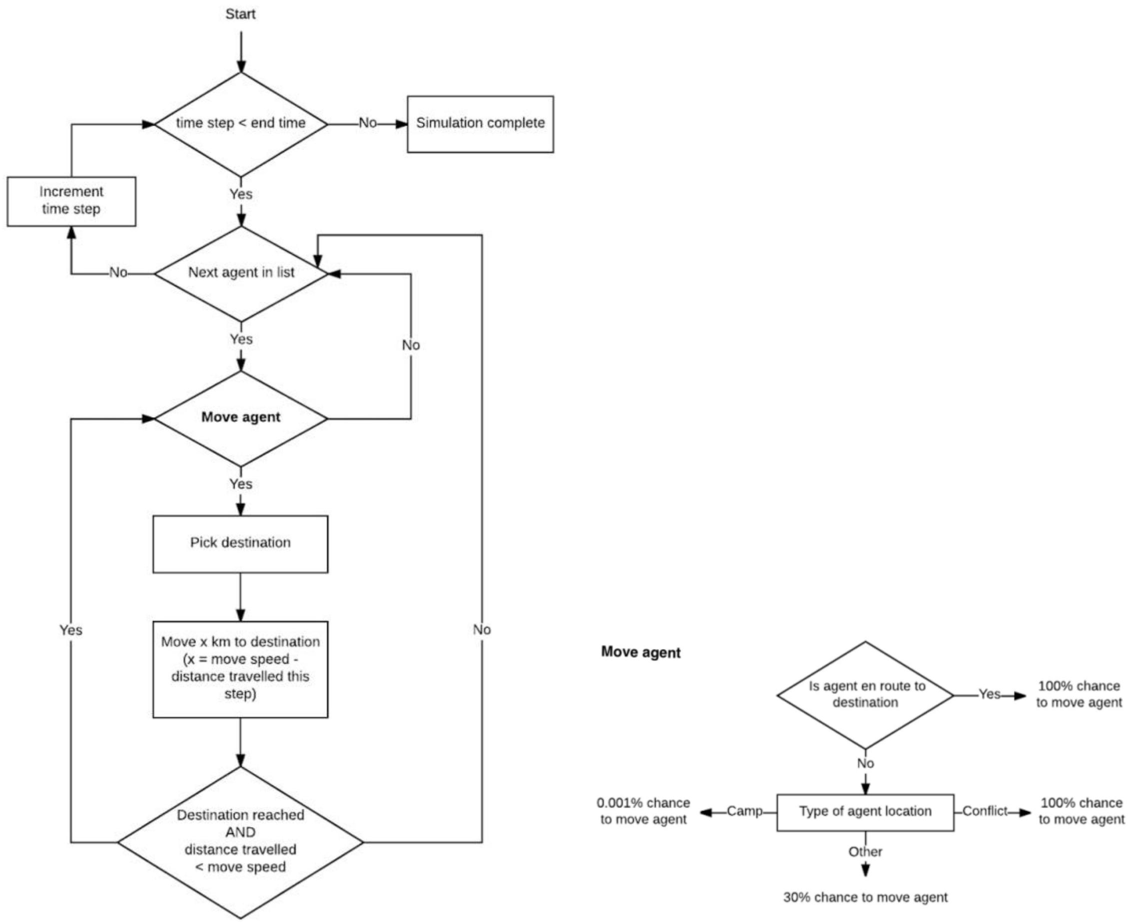


Fig. 5: Flowchart of algorithmic assumptions used in the Flee code to predict distribution of refugee arrivals along with 'move agent' factor. (Source: Suleimenova et al., 2017a, Supplementary Note 6).

5.2.3 Automation

The third foundational stage was automation of processes within the SDA (P3). Although formalisation of SDA helped provide a generalised framework for model construction, many processes within it were still manual. Automated model construction was required in order to enable quicker simulation construction for rapidly developing conflicts (Suleimenova et al., 2017b). The Flee code uses data extracted from publicly available databases, construction of network maps, and visualisation of outputs. These involve time consuming and manual steps that needed to be automated to respond to shorter time scales which conflict conditions demand. This required the development of an automated framework to enable the quick and systematic construction of refugee movements.

Automation helped improve the simulation process and integration of applications while creating the environment for users and researchers to curate and process data as well as construct models and modify simulations. Flee is automated using Fabric for Flee simulations (FabFlee). FabFlee is a combination of Fabric for Simulation (FabSim), a Python-based automation toolkit for simulation and data processing workflows currently in its third version (FabSim3, https://fabsim3.readthedocs.io/en/latest/) and the Flee simulation code. FabFlee works as a plugin for automated implementation of the SDA for forced displacement simulations. Figure 6 highlights the phases within SDA that stood to be improved with the integration of FabFlee.

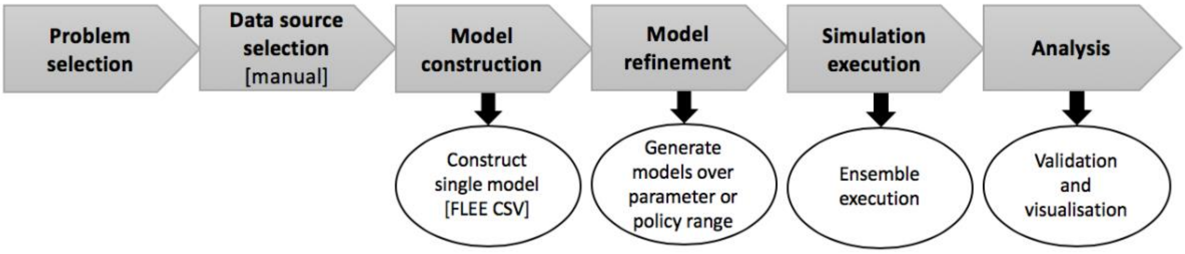


Fig. 6: SDA phases (in arrows) with automated FabFlee implementation (in ovals) from model construction to analysis. (Source: Suleimenova, 2020)

FabSim is a highly modifiable, general-purpose toolkit with multi-disciplinary applications which can automate a range of complex computational tasks (Groen et al., 2016). It has already been applied in diverse range of disciplines from cerebrovascular bloodflow to molecular dynamics. It helps automate routine administrative tasks and enables efficient management of complex computational problems, and effective use of distributed, remote computational infrastructures like HPC. Figure 7 provides an overview of the FabSim structure highlighting aspects of its flexibility and adaptability. It also provides the stepwise workflow on how it comes to be integrated within the model development process.

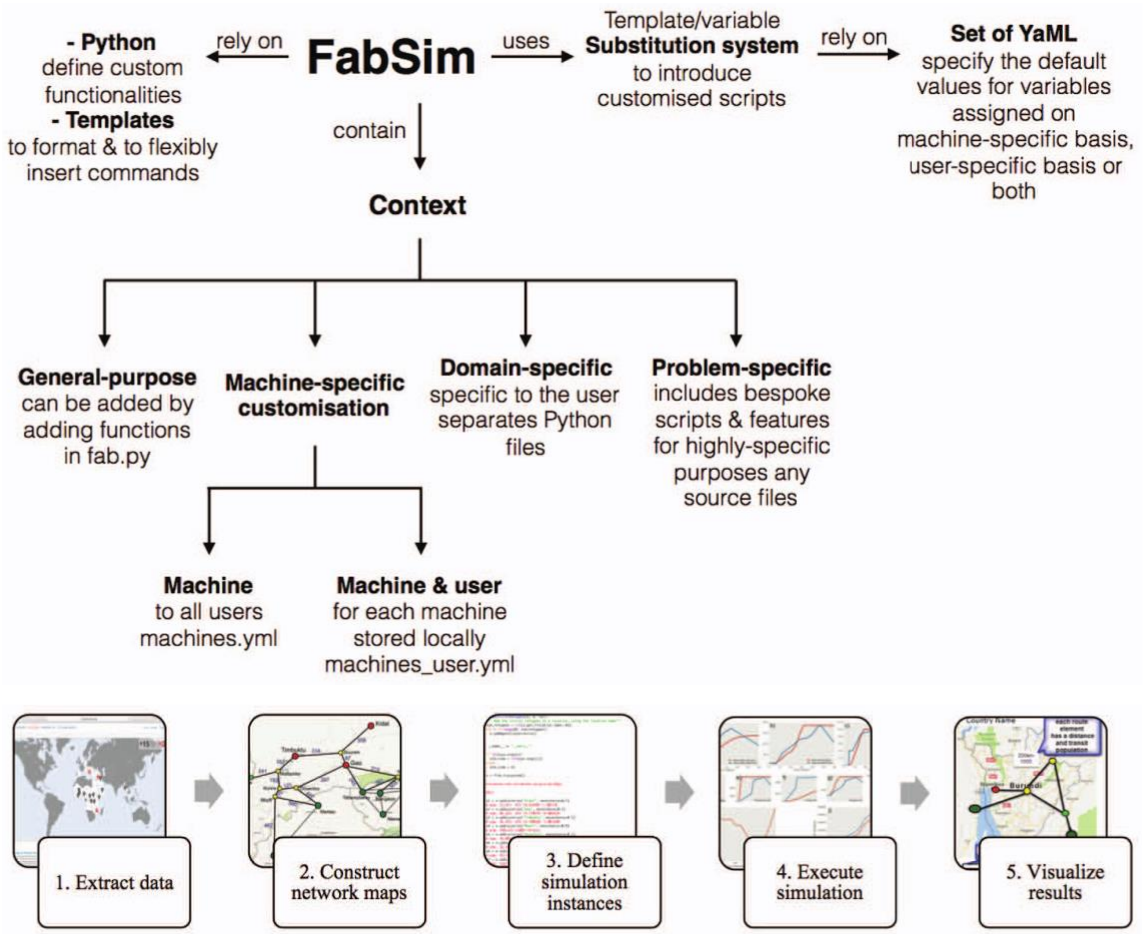


Fig. 7: FabSim structure and overview of steps required to construct, run, and visualise a basic refugee simulation. (Source: Suleimenova et al., 2017b)

The FabFlee automation toolkit helps optimise the Flee code to allow for easy adaptation to different conflicts, policy decisions, and assumptions. It helps in studying the effects of policy decisions like border closure, forced redirections, and speed changes etc. The basic workflow for FabFlee as illustrated in Figure 8 involves (1) loading the model with input data like location, routes, and conflict information; (2) adjusting simulation settings as desired; (3) instantiating the modified version of the conflict and running it to obtain results (Campos et al., 2019). FabFlee also helps in running multiple ensemble simulations and parallel execution of the code on HPCs as part of a software suite called VECMAtk aimed at integrated sensitivity analysis, parallel execution on supercomputers, and coupling (Suleimenova & Groen, 2020; Groen et al., 2021).

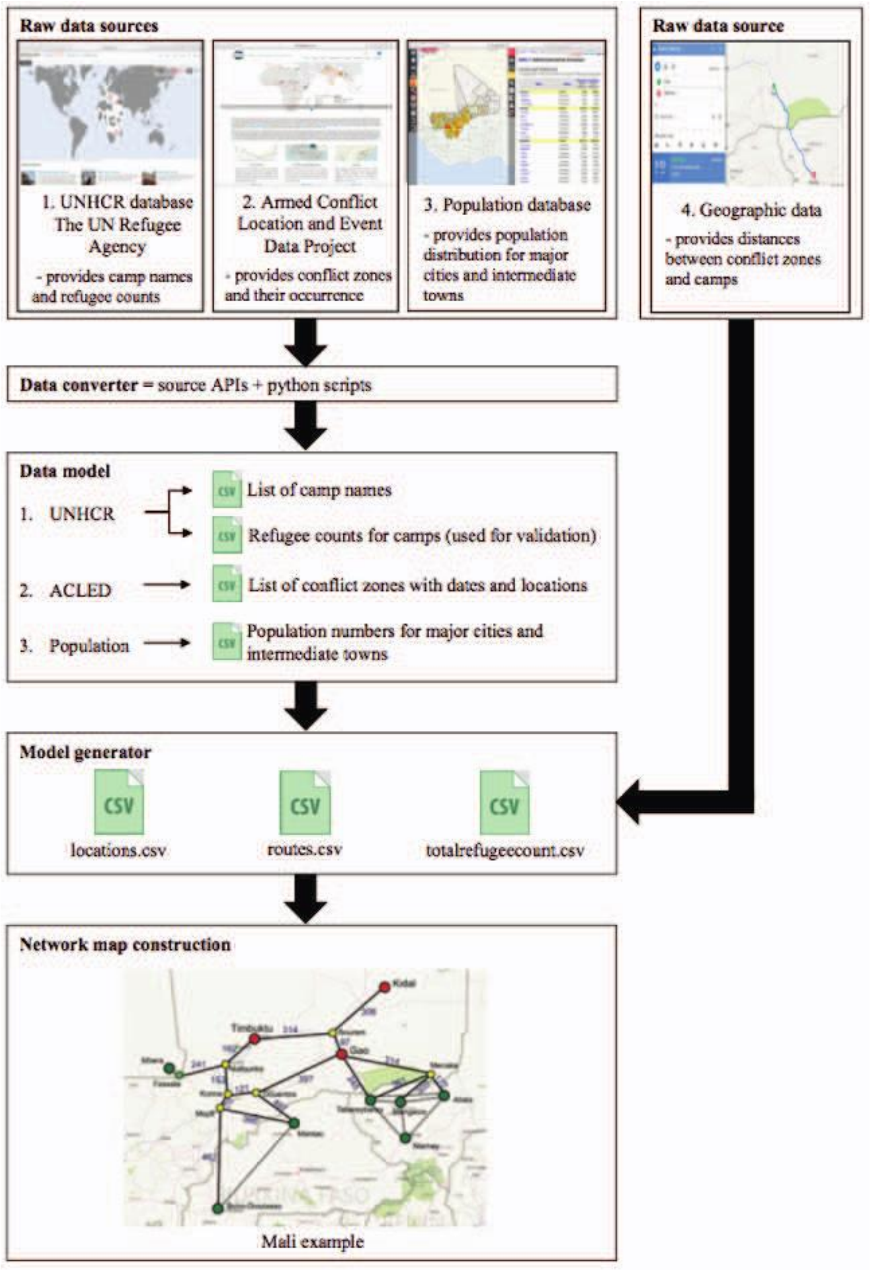
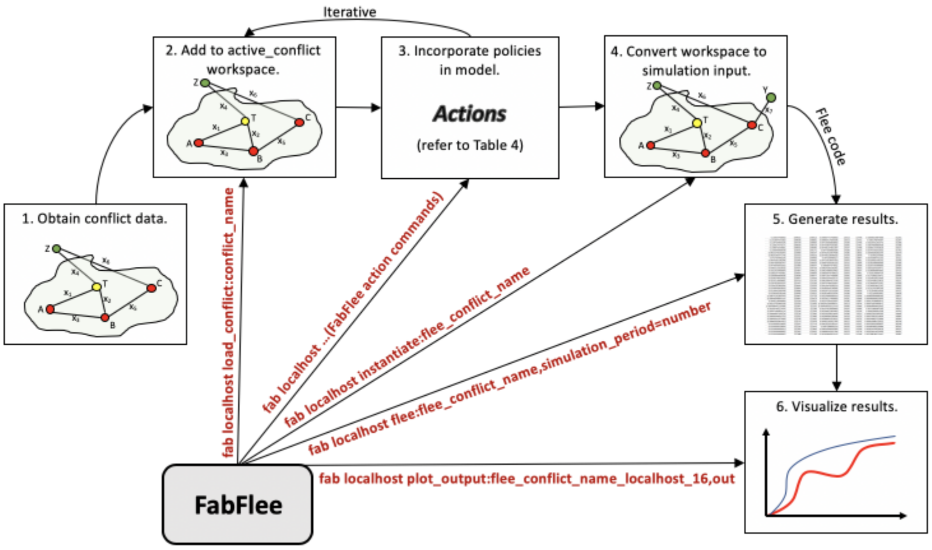


Fig. 8: Flee's model building approach (Source: Suleimenova et al., 2017b)

(a) Ensemble modelling of policy decisions

Apart from supporting a range of developments in Flee, FabFlee helped explore counterfactuals like the effect of different policy decisions like camp and border closures on simulation outcomes (P9). It also helped test the sensitivity of such outputs to different alternative policy decisions. Refugee camps are usually located in countries neighbouring the conflict country. An adverse policy decision by a neighbouring host country to close camps and borders could have a knock-on effect on refugee movement. Taking the conflict context of South Sudan and accounting for policy decisions like camps or border closures and forced redirections required an extension of the SDA (Suleimenova & Groen, 2020). As a part of the VECMA project this also involved presenting an automated policy exploration toolkit combined with sensitivity analysis. This included exploring systematic approaches to validate and analyse the sensitivity of simulations and to investigate variability in output due to sensitivity of parameters.

Incorporating policy changes in the SDA involves retaining its original six phases as described above but incorporating changes in policy decision in the refinement phase of the model by introducing both ensemble simulation executions and their sensitivity analysis. Ensemble simulation modelling involved refining the initial model with additional information and running the simulation many times, each time with different parameters and assumptions (Field Notes, Presentation delivered by DG for Oxford Brookes, 29/04/2022). Running the simulation with slightly different policy scenarios was aimed at highlighting how agents' decisions change in response to their awareness of different surroundings and how sensitive the output is to parameters like agents' awareness of possible routes and their speed of movement. FabFlee was used to automate and implement this ensemble modelling and different policy decisions were examined through their parameter explorations. This involved developing parameter exploration commands to modify according to the given range of parameters as highlighted in Figure 9 below. These included changes in camp capacity, new camp locations, removing an existing location, camp closure, border closure and forced redirection. These parameters were iteratively explored in steps 2 and 3 of the FabFlee workflow below to then generate and visualise results.



```
fab localhost load_conflict:conflict_name
fab localhost ...(FabFlee action commands)
fab localhost instantiate:flee_conflict_name
fab localhost flee:flee_conflict_name,simulation_period=number
fab localhost plot_output:flee_conflict_name_localhost_16,out
```

Actions	FabFlee command
change camp capacity	change_capacities:camp_name=capacity
add a new location	add_camp:camp_name,region,country,lat,lon
delete an existing location	delete_location:location_name
camp closure	close_camp:camp_name,country,closure_start,closure_end
border closure	close_border:country1,country2,closure_start,closure_end
forced redirection	redirect:source,destination,redirect_start,redirect_end

Fig. 9: FabFlee workflow diagram and steps for implementing different policy decisions as well functions for policy explorations. (Source: Suleimenova & Groen, 2020)

The simulation set up and execution was done by implementing the FabFlee workflow along with automatic sensitivity analysis for awareness level and movement speed for each policy scenario as illustrated in Figure 10 below.

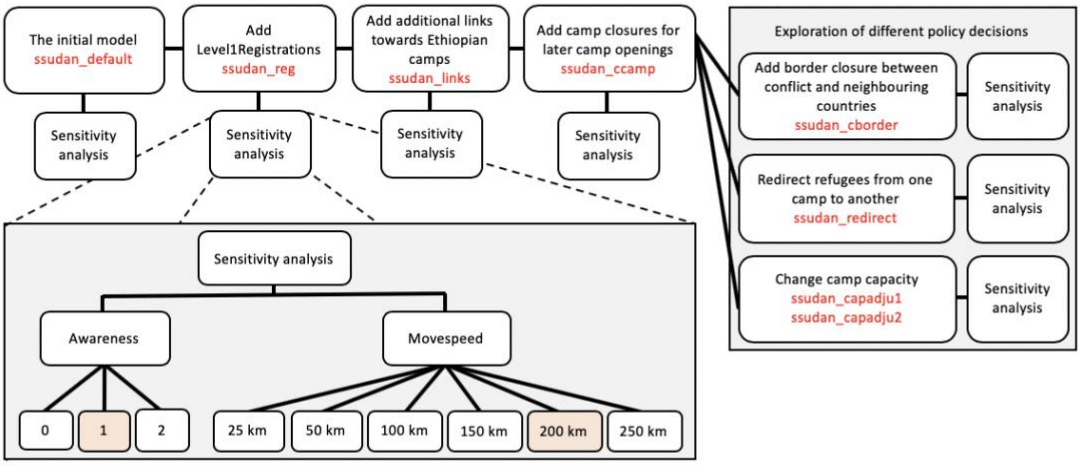


Fig. 10: Set-up for ensemble simulation execution for South Sudan (Source: Suleimenova & Groen, 2020)

For each ensemble run, sensitivity analysis was performed by varying the level of agent awareness and movement speed of the agents where awareness represents the level of a refugee agent's knowledge about nearby locations. Agents may know the path distance or type of location for adjacent locations (1 link away) or locations next to adjacent locations (2 links away) where policy decisions may influence the awareness of next available location and movement speed. Ensemble runs helped identify the effect of policy decisions on refugee movement by foregrounding emergent refugee behaviour like longer travel times for refugee agents and lingering effects of border and camps closures even after they have reopened. It also highlighted the need to explore terrain conditions when refugees have to take alternative routes in case of forced redirection and where terrain conditions affect the speed of movement. This instance also saw the incorporation of off-road links with modified assumptions of slower travel speed on the basis of a UNHCR report that stated refugees arrived in neighbouring Ethiopian camps on foot due to lack of roads.

5.2.4 Parallelisation, multiscale simulation and coupling

Groen (2018) developed the initial prototype of a multiscale simulation approach through the parallelisation of the Flee code and support for multiscale coupling (P5). Parallelisation of a code enables its simultaneous execution between local and remote supercomputers to evaluate and improve its performance, efficiency, and the scale of the model that could be executed. This resulted in the *pflee* algorithm for parallel implementation which was subsequently scaled in P11. The coupling approach paved the way for incorporating different causal conditions with the Flee code by creating and joining sub-models together with Flee or facilitating coupling of macroscale and microscale models as well as weather data (subsequently used in P7, P8, P10). The performance of the parallel version of the code was benchmarked on supercomputers like Hawk, Eagle, and Vulcan available within the HiDALGO consortium (P11). These benchmarking activities help flag time consuming functions within the Flee code that consumed computational overheads. For example, one such benchmarking activity helped highlight how Flee needs to be parallelised across agents as well as locations otherwise each parallel execution calls up locations graphs at each iteration clogging up computational processes (HiDALGO, 2019a). Identification of these errors helped evaluate performance for large and more detailed coupled simulations (HiDALGO, 2019b). Supportive activities like these helped drive the key functional phases identified by improving efficiency of the code.

(a) First instance of coupling – food security

Incorporating the effects of food insecurity on the movement of refugees within the Flee code was the first instance of coupling Flee – in this case with a food security sub-model (P7). This aimed to explore the relevance of food security on the modelling of forced displacement due to the combined effect of war and famine that was unfolding in South Sudan (Campos et al., 2019). To understand the effects of food security on refugee movement dynamics required an assessment of how it was related to the conflict. This was done by exploring linear correlation across time and space between the ACLED dataset and IPC (Integration Food Security Phase Classification) dataset (https://www.ipcinfo.org/). The IPC dataset provides food insecurity indexes with comprehensive population classification by country, region, and month according to the food security conditions. It identifies five food security phases ranging from 'minimal stress' to 'famine' and classifies populations of different regions according to these. When mathematical calculations showed food security to be relevant to the model, modifications to parameters were suggested so these can be extrapolated to similar model constructions for other conflicts. The modifications included making probability of movement dependent on the IPC index of each region at each time step to account for the fraction of the population estimated by IPC to be in a stress situation. This was to ensure that the fraction of the population affected by food insecurity leaves it's given location in an effort to bring the simulation closer to real refugee behaviour. The second modification involves the assumption that food insecurity is the only cause for the refugees to depart. This involves expanding the list of spawn locations to include both conflict locations and locations which the IPC index indicates to be stress locations. The weighted probability of agents spawned in the conflict location involves the entire population according to the last census and in the IPC locations that are not conflict zones, fraction of the population experiencing stressor according to IPC.

These modifications were incorporated through a separate sub-model within the original Flee code that either updates the movement probabilities of neutral location or modifies the spawning probabilities of refugees over time. It starts with an input data file that contains the IPC indexes which enables the food security sub-model to update parameters in affected locations in tandem with changes in IPC over time. This also involved implementing new commands in FabFlee to run the simulation while including the food security sub-model and generating side-by-side comparison graphs of key metrics between both types of simulation.

FabFlee enabled multiple different simulations to be performed for the same conflict without manually editing every part of the code each time. This helped in studying the effects of food security on refugee movements through a sub-model, by adding a few modules to FabFlee to account for these modifications, without having to construct a simulation model from scratch. This enabled both models to run under the same general conditions i.e. with and without the modifications.

(b) Second instance of coupling – Flare conflict evolution sub-model

To model the effects of violent conflict evolution the Flee code was coupled with a sub-model called Flare which aims to forecast where violent events are likely to occur next based on historical occurrence (Groen et al., 2019a) (P8). This instance of coupling used a hybrid simulation approach to combine Flare's stochastic network-based algorithm with Flee's ABM. The approach was implemented in the case of the 2012 Mali conflict and was another step in moving towards enhancing the accuracy of a particular model. This involved exploring multiple causal relationships like electoral or political violence as well as the relationship between mining and mineral prices with mining activity and spike in mineral prices increasing the risk of conflict in areas producing such minerals. While the SDA relied on historical conflict data, forecasting refugee arrivals as a result of *future* conflict required the integration of a conflict evolution model which estimates how violence in a civil war evolves over time. Incorporating this required modification of the SDA and replacing some activities in the data source selection and model construction phases.

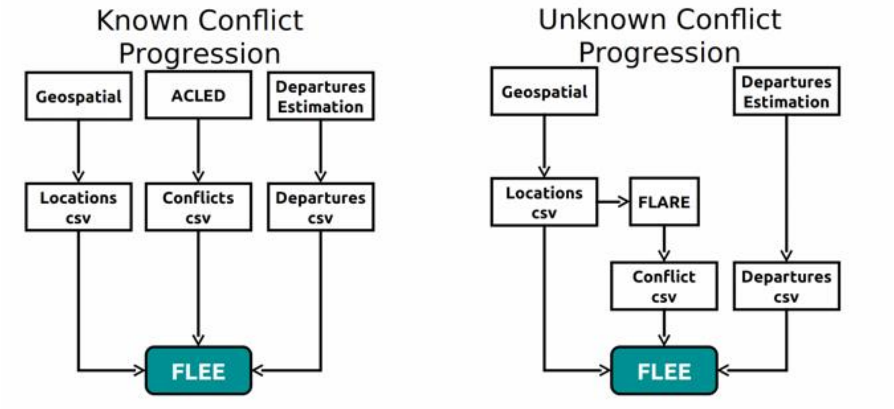


Fig. 11: Overview of SDA modification required to incorporate conflict evolutions (Source: Groen et al., 2019a)

As illustrated in Figure 11 above, in contrast to known conflict progressions where ACLED provided the conflict input data for the Flee model, forecasting refugee arrivals involved using the Flare model's output data instead. The Flare code requires three inputs in the form of a list of location, routes that connect them (geospatial data), and an initial location where conflict is known to have occurred. Using these data inputs it predicts the potential locations to be affected by conflict and the period during which such conflict is likely to occur. This information is then written into the conflict input file which feeds Flee. During each iteration the Flee code checks for this information to determine whether a given location is a conflict zone and modifies appropriate movement probabilities and location attractiveness values if and when the status of given location changes. To enable Flee's conflict information processing, adaptation was required to its data manager to be able to handle an additional file format in the form of a conflict input file.

Conflict evolution can be a multicausal phenomenon, and for coupling with the Flare sub-model, a particular cause was identified which in this case was mineral mining and pricing, particularly because the role of mineral mining and pricing could have a stronger effect on conflict evolution in mineral rich countries like Mali. Towards this end, ACLED data on conflict occurrence was compared with the MSCI Worlds Metal and Mining Index which showed the indication of a potential relationship to understand locations where potential conflict can erupt.

(c) Third instance of coupling - Multiscale simulation and weather coupling

Multiscale modelling involves the combination of models at different temporal and spatial scales (Groen, 2018) (P10). Combining models at different scales was a recognition of the complex causality inherent in forced displacement which would help increase predictive and forecasting accuracy (Jahani et al., 2021). Using South Sudan as the test scenario, this involved defining a macroscale model for most of South Sudan during a given conflict period and a microscale model that covers the region around White Nile which in turn was coupled with weather data like precipitation and river discharge to determine how they affect refugee movement. South Sudan has alternating wet and dry climate with floods and droughts being the most significant natural disasters as a result of which precipitation and river discharge were incorporated as causal factors within the model. This included weather data provided by European Centre for Medium-Range Weather Forecasts (ECMWF), a consortium partner within the HiDALGO project. This included precipitation data from its Climate Data Store and river discharge data from the Global Flood Forecasting System (GloFAS) provided by the Copernicus Project which is the earth observation project of the European Union Space Programme of which ECMWF is also a partner.

The model included 8 regions in South Sudan and 14 camps in 4 neighbouring countries - Uganda, Kenya, Sudan, and Democratic Republic of Congo. The proposed multiscale prototype divides the whole model into two sub-models which are executed independently of each other with agents allowed to pass between the models as they run for the same conflict period. Each location in which agents pass between models is known as a coupled location. In addition, all locations from the microscale model are added to the macroscale one without any links to the locations in the latter and are known as ghost locations. This is to facilitate the macroscale model in inserting agents into these locations according to the Flee algorithm and for the coupling interface to then transfer all agents from each ghost location into the microscale model at each time step.

The microscale models aim to capture key walking routes, roads, and river crossings in the mountainous areas in South Sudan. Coupling these with weather conditions increased the level of detail within this model. The microscale model aims to predict forced migration movements from the Upper Nile and Jonglei towards camps in Gambela, Ethiopia with additional algorithmic assumptions involving drive, walk, and river routes that affect movement speed in the region. Once the macroscale and microscale models are individually constructed they are linked through two coupling approaches – file coupling and model coupling using MUSCLE 3 (Multiscale Coupling Library and Environment version 3). These two coupling approaches are used to evaluate their comparative performance across scenarios. Along with the micro-macro coupling, the microscale model is coupled with the weather data. The simulations were run based on Flee rule set 2.0 which was a work-in-progress in 2019 and formalised in Suleimenova et al. (2021) through the Sensitivity Driven Simulation Development (SDSD) approach. (Internal document on Flee 2.0 rule set; first review feedback). Flee 2.0 essentially made changes in parameters and assumptions based on conversations with humanitarian organisations, NGOs, and researchers in the field (Internal document on Flee 2.0 rule set).

5.2.5 Sensitivity driven simulation development

Sensitivity analysis had been performed in some form over the course of Flee's development but was integrated with the SDA and formalised through the SDSD approach through the rule set update in the form of Flee 2.0 (P13). The use of sensitivity analysis guides further simulation development and refinement efforts without the need to directly calibrate with validation data (Suleimenova et al., 2021). This is because certain parameters like movement speeds and awareness levels were weak assumptions based on intuition and interaction with humanitarian organisations and the feedback received from them introduced a level of nuance (Internal document on Flee 2.0 rule set). Sensitivity analysis helps identify assumptions to which validation results are particularly sensitive to. This then helps in refinement of the model's rule set and in balancing sensitivity more evenly across different assumptions and parameters.

Sensitivity of the output to relatively trivial elements in the system indicates that the model is not accurately balancing the main influencing factors which would help in achieving greater model accuracy. Given an existing simulation, SDSD proceeds through four steps as illustrated in Figure 12 below: (1) using sensitivity analysis techniques to measure the sensitivity of key assumptions in the simulation; (2) sensitivity analysis helps highlight the parameters that have a large or disproportionate impact on the simulation output, these are labelled as 'pivotal parameters'; (3) Examining the underlying model logic involving these parameters and manually extending the model and implementation. This can be done by adding further details in the simulation of such parameters like adding additional rules, detailed breakdown, or incorporating derivative parameters; (4) In light of the above three steps, the last step involves evaluating whether the simulation is fit for purpose or repeating the whole process again.

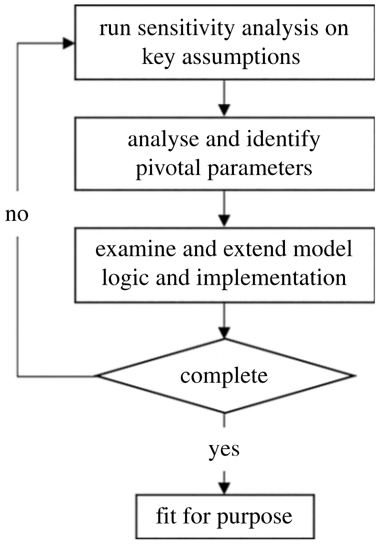


Fig. 12: SDSD approach flowchart (Source: Suleimenova et al., 2021)

The sensitivity analysis was performed using a combination of software tools which form a part of the VECMA toolkit (VECMAtk: https://www.vecma-toolkit.eu/toolkit/). These include EasyVVUQ, QCGPilotJob, and FabFlee in combination with the Flee code. EasyVVUQ (https://github.com/UCL-CCS/EasyVVUQ; https://easyvvuq.readthedocs.io/en/dev/) helps validation, verification, and uncertainty quantification for a wide variety of simulations. QCG-PilotJob (https://github.com/psnc-qcg/QCG-PilotJob; https://qcg-pilotjob.readthedocs.io/en/develop/) helps in the execution of multiple computational tasks within a single job allocation slot on supercomputers. This is important because supercomputers normally restrict the number of job allocation slots that can be used by a user at any given time (DG second review feedback).

Based on sensitivity analysis of the existing parameters within Flee two modifications were implemented: (1) Introduction of the Recent Distance Travelled Index (RTDI) based on the assumption that if an agent has travelled a certain distance, they would benefit from a break. The RTDI can have threshold value that is between 0-1. If the RTDI threshold is set to 0.5, then when the agents set out they will travel at least for a day at maximum speed before requiring a break. If they travel less than that they will continue to move i.e. a higher RTDI threshold will lead to the agents travelling longer without breaks while a lower one would introduce more frequent breaks. (2) Maximum movement parameter was refined by proposing a new range value and adding a new mode of transport i.e. walking based on qualitative research conducted with NGOs on field.

Qualitative research showed that people travel 3-4 kms on foot when they originally depart due to roads being blocked by armed forces and not having secured proper shared transportation by then since there is evidence that people used shared vehicles. Moreover, the vehicles may not arrive immediately, need to make detours and stops to transit other people. This translates to an average walking speed of 30-40km per day with an average of 12 hours per day. Based on these insights the parameter on movement speed was pegged to vehicular travel only where the range was refined to 100-420 km from 20-200 km with a new additional parameter on walking speed set to a maximum of 35 km per day. However, the average speed could not be fixed in any of the simulations despite refinement due to lack of validation data, similarly for probability of refugees moving from camps, despite it being a pivotal parameter, due to lack of data on camp conditions. While sensitivity analysis helps in guiding model refinement, iterative sensitivity analysis can be computationally expensive (particularly in case of large number of assumptions with varying dimensions) though in this case they were managed through supercomputers and relative simplicity of the Flee code. It also requires an in-depth understanding of the simulated problem because the value of the sensitivity analysis lies in the implementation of refinements based on insights which cannot be automated. This highlights the time taking aspects of familiarising oneself with different aspects and dynamics of particular conflicts.

5.2.6 Simulation – Optimisation Approach

The aim of the simulation – optimisation approach was to combine the Flee ABM with a machine learning based evolutionary algorithm to address the issue of camp placements as a MOO problem (Xue et al., 2022) (P14). This was also the first instance of using the ITFLOWS recommended terminology of ‘asylum seekers / unrecognised refugees’ to homogenise its use across implementations. Camps are important infrastructures for aid delivery and deployment for individuals fleeing violent conflict and humanitarian organisations face significant challenges in determining optimal camp locations. The issue of camp placement was formulated as a Facility Location Problem (FLP), first explored in P4, and as a MOO problem. It was resolved through the optimisation of three objectives: minimisation of travel distance, maximisation of camp arrivals, and minimisation of idle camp capacity. This was executed through the combination of Flee ABM model and NSGA-II evolutionary algorithm using the route pruning algorithm developed in collaboration with KNOW Centre which helped speed up visualisation through automated construction of location graphs at scale (P12). The approach works with the NSGA-II algorithm generating candidate locations with the Flee algorithm taking the coordinates of these locations as an input to generate the simulation output for the optimisation stage. These are then evaluated on the basis of an indexed evaluation of the number of camps to be placed and opened. Within this prototype version, the number of camps to be opened were taken to be 1 in order to reduce computational cost and run-time. The simulation was automated using FabFlee and scheduled on a supercomputer using QCGPilotJob.

5.2.7 Integrated model – Tigray

A request from Save the Children to model the then ongoing Tigray conflict (P15) in 2020 and the subsequent collaboration resulted in forecasting forced displacement and modelling IDPs (Suleimenova et al., 2022, Interview with DS, 15/07/2021). The collaboration had initially started with a focus on Burundi but the focus shifted to Tigray with the exacerbation of the conflict in the area (Interview with DS, 15/07/2021).

> *In September 2020, ... Save the Children contacted us. They sent an email saying we have read your paper and we would like to ask for a collaboration with you. So after that we had a call with them and they described what they do and we described what we do – as first meetings go. And then they said, can you actually develop a model for Burundi. So they were interested back then on Burundi. So I started working on that and then we had another call and they said 'hey, do you know what happened?' The conflict in Tigray had just started. That was two months later in November. 'So what do you think of developing and forecasting it?' [they asked] Before we had done only predictions. By predictions we mean historical events that happened so that we can compare our results [with existing datasets]. But this time they asked for forecasts. [Interview with DS, 15/07/2021].*

This was a work-in-progress at the time of field work and the product of ongoing experimentation and testing through consultation and feedback not just with Save the Children but also with IOM on the feasibility of modelling IDPs due to lack of access to data in real time and the difficulty in pegging down parameters to define an IDP rule set (Field Notes). This is because IDPs have different push and pull factors compared to refugees. However, the exercise was undertaken within the context of the unfolding conflict because IDPs and Eritrean refugees within that region could become refugees fleeing to Sudan with the intensification of the conflict.

This involved review and revision of key assumptions within the Flee code and searching and incorporating new information with the aim of forecasting movement patterns, destination preferences, emerging trends for destinations within Sudan. Like the initial Flee model this started with simple assumptions in conjunction with a guided randomizer for different conflict scenarios using which multiple real time forecasts were made during active collaboration with Save the Children which were then refined through feedback (Internal file on forecasting the Ethiopian conflict for Save the Children).

The simulation included IDPs and Eritrean refugees in the Tigray and not in other regions of Ethiopia due to lack of data. Both refugees and IDPs are taken into consideration because IDPs have the potential of becoming refugees if the conflict intensifies. Further, it was assumed that displaced people travel on foot at a speed of up to 40 kms per day on the basis of insights gained through the collaboration. This allowed agents to depart from their location up to 2 days before the eruption of the actual conflict due to spread of danger warnings from nearby locations or families and friends. Conflicts were introduced at random locations in selected regions of Tigray and their frequency was determined by the chosen level of intensity which allowed the simulation of agent movements on the basis of different conflict scenarios. 18 different conflict scenarios were defined in 5 potential districts with three levels of conflict intensity (low with 5 events, medium with 10 events, and high with 15 events). Each conflict event was assumed to take place in one location and last between 2 and 20 days.

While the Flare code demonstrated the first attempt at forecasting future events, it aimed to forecast potential conflict events based on past data while this approach aimed to forecast refugee distribution based on different conflict scenarios. Since it was a work in progress, limited conclusions could be drawn from the model though the model showed that people fleeing conflicts further away from the camps in Sudan tend to go to Eritrea, Djibouti, or other parts of Tigray in the short term but may eventually arrive at Sudan. This tended to have a more gradual boosting effect over time while nearby conflicts saw an immediate spike in arrivals.

5.2.8 Flee 3.0

Flee 3.0 (https://github.com/djgroen/flee/releases/tag/v3.0) represents the third rule set update for Flee taking into account insights from the collaboration with Save the Children as well findings of ITFLOWS deliverable 3.4 on *Time sequence of forced displacement into neighbouring countries* (ITFLOWS, 2022; Internal documents; Field notes). This involved providing support for demographic attributes of agents and adding new demographic-based movement rules. This was based on the insight that it would be useful to tailor support based on demographic attributes like children or elderly. This was also driven by the insight that displaced persons demonstrated ethnic preferences in moving towards particular camps i.e. moving towards camps of their own ethnicity. It also incorporated support for conflict driven agent spawning where earlier agents were spawned on the basis of input data fed to the model by the user. It further incorporated support on IDP camp locations allowing them to be opened or closed and the ability to vary conflict intensity which in the earlier versions were hard coded. Flee 3.0 was a work in progress and there was discussion on incorporating a graded post-conflict dispersion of displacement persons with more people being displaced in the first few months after outbreak with a gradual outflow of people over time as conditions can be exacerbated due to ongoing conflict and climate change (ITFLOWS, 2022). It also highlighted the need to capture the more gradual mechanism driving displacement in conflict zones as a combination of food insecurity, economic conditions, and persistent instability and tensions (Field notes; Team meeting #53). Flee 3.0 was a nascent work-in-progress at the time of concluding field work.

5.3 Ecosystem: The core team, stakeholders, collaborations, and projects

As mentioned earlier, the first version of Flee was developed as a prototype by DG as a single developer with simple parameters and assumptions to illustrate the possibility of simulating a complex and multicausal social phenomenon such as forced migration. DS joining as a PhD student in the Department helped advance work towards generalisation and automated evolution of Flee. A substantive part of Flee’s foundational development process forms DS’ PhD thesis covering generalisation (SDA), automation (FabFlee), and refinement (policy decisions) of the Flee code (Suleimenova, 2020). After her PhD she continued with the team as a Postdoctoral Researcher working on VVUQ, sensitivity analysis, and building new conflict models – particularly modelling IDPs and scenarios for the Tigray conflict (Field notes). The core team responsible for the primary development of the Flee code expanded over time with the joining of HA who made significant contributions to FabFlee development, API integration, and seamless integration and deployment of Flee and its suite of software tools across different computational environments and systems. AJ joined as a Postdoctoral researcher in 2020 and extended the work on coupling, multiscale migration modelling, while also working on new and existing conflict models particularly for the ITLOWS project (Field notes). YX joined as a Postdoctoral researcher in 2021 working on MOO problems for camp location placement towards an integrated simulation-optimisation approach drawing on her expertise in machine learning algorithms (Field Notes; Xue et al., 2022).

Expanding the team over time was possible due to funded research projects and scientific consortia. In the beginning, its development was sustained as a PhD project with support for DG's time coming from the ComPat grant that carried over from his time at the UCL. During this time, the nascent development was also supported through collaborations with existing academic networks and workshop participations to explore new scientific questions. These existing academic networks were then sustained through funded research projects. VECMA and SEAVEA was anchored by DG's former colleagues at the UCL while ITFLOWS is jointly led by the BUL Department of Law. DG's former colleague from the University of Amsterdam introduced him to the CoeGSS who were putting together the proposal for HiDALGO. As highlighted above, this was at a time when DG did not have any other grant and was trying to get some form of support. Around the same time, he met representatives from UNHCR at an event following which they both shared their ongoing work in the area of predictive analytics for forced displacement and collaboration ensued. As a result of this collaboration, a letter of support was provided by the UNHCR for the HiDALGO funding application. As these existing networks materialised through funded consortia, they not only provided the financial resources to expand and sustain the core team but also provided access to resources like supercomputers and diverse range of expertise among consortia partners. They also shaped the development of Flee through project mandates and cross-collaboration across consortia, thereby supporting complementary workstreams. Both VECMA and ITFLOWS were associate partners in HiDALGO. HiDALGO consortium partner ECMWF was an associate partner in VECMA while PSNC, a consortium partner in HiDALGO, was also a consortium partner in VECMA and SEAVEA. Associate partnerships represented external collaborations beyond the funded consortia and helped to leverage complementary research. Apart from funded projects, there were collaborations with humanitarian organisations and academic research networks that have had a significant influence in critical areas of Flee's development. The following sections outline key projects and relevant partnerships and their role within the development process:

5.3.1 Funded projects

(1) HiDALGO (https://hidalgo-project.eu/): It was funded under the European Union's Horizon 2020 Framework and included members from 13 institutions across 7 countries. The aim of the HiDALGO project was to develop advanced computational techniques and simulation frameworks in answer to some of the world's biggest challenges. The project was structured into 8 work packages dedicated to managerial, technical, and knowledge exchange work streams. BUL was actively involved in the technical work packages on HPC and HPDA System Support and Pilot Applications. Out of the consortium partners, its closest working relationships were with ECMWF, KNOW Centre, PSNC (Poznan Supercomputing and Networking Center), HLRS (High Performance Computing Center Stuttgart), and National Technical University of Athens in collaboration with whom they made significant advances and additions to Flee. PSNC hosted the supercomputer Eagle and Altair and provided service support to running simulations on supercomputers while HLRS hosted the supercomputer Hazelhen, Hawk, and Vulcan. Supercomputers were used in P8, P10, P11, and P13 for coupling, scaling, and sensitivity analysis. However, Iranian Postdoctoral researchers within the team were unable to use supercomputers at HLRS because they were made in the United States and therefore fell within legal sanctions imposed by the United States on certain nationalities (DG second review feedback). Along with the National Technical University of Athens, they advanced the parallelisation of Flee for its efficient execution on supercomputers (Anastasiadis et al., 2021; HiDALGO, 2019a; 2019b; 2021a) (P11).

The multiscale modelling project with macro-micro coupling and weather data coupling were undertaken in collaboration with ECMWF (P10). ECMWF is an independent intergovernmental organisation supported by 35 member states. It serves as both a research institute and a 24x7 operational service (https://www.ecmwf.int/) which produces global weather predictions and other data for their member and co-operating states as well as the broader community. They are also a key consortium partner in the Copernicus project (https://www.copernicus.eu/en) of the European Union which serves as the Earth Observation component of the EU Space Programme offering quality assured data in the domain of atmosphere, marine, land, climate change, security and emergency. ECMWF's role within the consortium was a part of their long term modernisation programme to develop forecast models and product chains for exascale computing through collaboration with hardware vendors, research centres, and universities. The multiscale modelling project (P10) included data provided by the ECMWF and GloFAS early warning system from the Copernicus project's Emergency Management System. ECMWF's role was to help users build downstream application using the data available with them and they contributed their expertise not only during the South Sudan multiscale and weather coupling but also during the Mali route and terrain accessibility student dissertation project located at the University of Utrecht with co-supervision from a member of the Dutch Ministry of Defence who was also a former geospatial analyst (Field notes).

Within the consortium, the KNOW Centre offered competencies in the areas of artificial intelligence, deep learning, and other data science methods to drive workflow implementation as well as visualisation efforts. KNOW Centre helped develop a route pruning algorithm for automated construction of geographic location graphs (Schweimer et al., 2021) (P12). The automated construction of location graphs was aimed at circumventing the time consuming and error-prone procedure of constructing manual location graphs. The route pruning algorithm was applied in the optimisation-simulation approach to calculate the distance between the camp and its nearest location within South Sudan (Xue et al., 2022) (P14).

An intended collaboration with Moonstar to couple telecommunications data with Flee did not work out due to lack of utility of the data provided for ABM migration modelling. Moonstar is global network service provider, headquartered in Germany, whose role within the project was to provide anonymised telecommunications data to pilot applications. Within the Flee migration pilot, the purpose of coupling with telecommunications data was to use this data to model the journey of refugees from conflict zones to camps, particularly within the context of the Syrian conflict. However, the data only provided limited information like call duration or the number of calls made each day. It did not provide the spatio-temporal distribution of the phones of anonymised users because of which it could not be coupled with Flee code (HiDALGO, 2021c).

(2) ITFLOWS (https://www.itflows.eu/): It is an interdisciplinary consortia at the intersection of migration, technology, and human rights that aims to provide predictions of migration flow to the European Union to facilitate reception, relocation, settlement, and integration. Like HiDALGO it is also funded by the European Union's Horizon 2020 Research and Innovation programme. The primary objective of the project was to develop a predictive tool (the EUMigraTool or EMT) to help municipalities and humanitarian organisations pre-position resources and support for arrivals. While the project had a wider scope of also identifying potential sentiments and tensions between people residing in the EU and new arrivals, Flee as a part of the EMT was to be used as an input for prediction support. EMT comprised of two complementary approaches that included Flee and a large scale model. Flee as a small scale model provided predictions on camp arrivals in countries neighbouring the conflict country. This served as an input for the large-scale model that aimed to produce monthly prediction of asylum applications in the EU using machine learning approaches like neural network architectures and time series analyses while also allowing for correlation analyses between raw data and simulations.

Within ITFLOWS, Flee was required to develop scenarios for five conflict situations Afghanistan, Mali, Syria, Nigeria, and Venezuela. The selection were based on the rates of asylum application to the EU according to countries of origin. While the Mali conflict was quite well developed within Flee, the rest of the conflict scenarios had to be constructed from scratch. This involved significant secondary research into understanding the conflict dynamics through research and news reports that could inform the modelling process. For example, political and economic instability in Venezuela from the mid-2010s made Colombia the main transit and destination country for Venezuelan migrants and asylum seekers. Within EU, Spain was the main destination country due to historical and language ties (ITFLOWS, 2021d). For the modelling process, this presented particular complications - since the crisis in Venezuela was not conflict driven there were no flashpoint events that triggered population movement but more gradual socio-economic mechanisms which were difficult to incorporate within the existing modelling process. Further, with countries like Afghanistan there were no UNHCR offices in neighbouring Iran even though Eastern Iran has shared language ties with parts of Afghanistan (Field notes) and overall UNHCR numbers were thought to be outdated for the country after several months of efforts (Field notes). There were significant difficulties in developing a credible model for Afghanistan (at one point in a team meeting after almost 8 months, AJ remarked *"we have nothing, literally nothing for the data"* (Team meeting #42). This highlighted the difficulties in constructing input files for Afghanistan to run Flee. Therefore, with the outbreak of war in Ukraine, a decision was taken to replace Afghanistan with Ukraine. The need to model Ukraine was also echoed by other partners like Colombia University (Field notes).

Due to the interdisciplinary nature of the consortia there was significant emphasis on legal harmonisation and ethical considerations. For example, Flee used data sources from the UNHCR whereas the output of the EMT is intended for being used by institutions within the EU and definitions and thresholds are different for who is considered a refugee as opposed to an asylum seeker. Given the differing definitions within different datasets used by different groups within the consortia, this placed restrictions on what could be predicted using those datasets as well as concerns about how the outputs resulting from them would be used. This required arriving at a terminology that could straddle these differences while still retaining usability of the output within the overall objectives of the project. As a result, a mapping exercise of different definitions across datasets was undertaken and the term 'asylum seekers / unrecognised refugees' were put forward and which came into use in P14 in the development of the simulation – optimisation approach.

(3) VECMA (https://www.vecma.eu/) and SEAVEA (https://www.seavea-project.org/): The VECMA project aimed to support a diverse set of multiscale applications to run on exascale supercomputing environments with high fidelity so that their outputs are actionable. In other words, the aim was to make outputs from computational models reliable for real world application by ensuring such outputs are demonstrably validated, verified, and uncertainty quantified through appropriate mathematical techniques. The main deliverable was an open source toolkit containing a suite of software that would enable VVUQ to be performed on multiscale models. The VECMA project provided the opportunity to test the sensitivity of parameters to the output and also helped integrate sensitivity analysis with the SDA through formalisation of Flee rule set 2.0. The VECMA toolkit or VECMAtk comprised of a number of software including the ones used by Flee like FabSim3 on which FabFlee is based, EasyVVUQ that facilitates integrated sensitivity analysis, QCGPilotJob that facilitates scheduling execution requests on supercomputers, and MUSCLE 3 that provides the

computational environment that enables coupling. Tools like FabSim 3 and QCGPilotJob are extensively used for Flee within the HiDALGO project as well as in the modelling of the Tigray conflict that was being done in collaboration with Save the Children (VECMA, 2021b). The combination of FabFlee and EasyVVUQ helped perform sensitivity analysis for parameters like agent awareness, walking speeds, and probability of movement from given locations (VECMA, 2019). This also allowed automated parameter exploration and uncertainty quantification for inputs. Uncertainty quantification and sensitivity analysis in simulations help understand the scenarios in which the model performs well. These helped in performing sensitivity analysis on four African conflicts leading to the formalisation of Flee 2.0. Even though the conflict simulator Flare was developed in collaboration with the HiDALGO project (VECMA, 2019), it was also tested for uncertainty quantification within VECMA. The SEAVEA project is a successor grant that essentially continues the development and maintenance of VECMAtk (as SEAVEAtk) and aims to contribute to the UK ExCALIBUR (Exascale Computing Algorithms and Infrastructures Benefiting UK Research, https://excalibur.ac.uk/) community. ExCALIBUR is a UK research programme that aims to deliver high performance simulation software for high-priority research areas in the UK. Unlike VECMA which was funded under the EU Horizon 2020 programme, SEAVEA was funded by the UKRI's (UK Research and Innovation's) Engineering and Physical Sciences Research Council (EPSRC).

5.3.2 Humanitarian organisations

During the course of its development, the team came in contact with multiple humanitarian organisations. However, the relationships were often ad-hoc, serendipitous, or in response to an urgent requirement.

> DG mentions in a meeting that working in the Computer Science department they have tools but someone needs to use it. That the hope is that these are used to solve relevant issues rather than remaining as purely academic endeavours. They remain keen to collaborate with NGOs but in the past there have been some hits and misses. Some have approached and lost interest. But the collaboration with Save the Children sustained as they were willing to work together to see what works, acknowledging that it is a time taking process (Field notes - Humanitarian Organisations #4).

The first organisation that they came in contact with was UNHCR whom DG had met at an event (Interview with DG, 18/10/2021). They subsequently got in touch for knowledge exchange since at the time they themselves were developing a model for forced displacement based on machine learning and were trying to understand the landscape in terms of what else is out there. Flee was one of the existing models with proof of concept. During subsequent discussions and knowledge exchange UNHCR suggested that the team focus on South Sudan while they worked on Somalia and for a while both developed models parallelly (Interview with DS, 15/07/2021). As mentioned earlier, UNHCR also wrote a letter of support for Flee for the HiDALGO funding application (Interview with DG, 18/10/2021).

The collaboration with Save the Children came about in September 2020 when they reached out to the team after reading one of their papers (Interview with DS, 15/07/2021). Initially they requested development of a model for Burundi. However, while that model was being developed, in two months' time, they reached out again in the backdrop of the outbreak of conflict in Tigray to request a shift of focus so that reports and insights could be distributed to NGOs on the ground in Sudan and Ethiopia. This was the first time that Flee was going to be used for forecasting i.e. making projections into future as well as for modelling IDPs and that too for a conflict unfolding in real time. These efforts culminated in P15 for integrated refugee – IDP forecasting. Prior to this, Flee had only simulated past conflicts and validated the results against existing data from UNHCR. As a result of this collaboration, they first constructed a model and validated it for a month to understand if they were going in the right direction. This model was then run for the next three months on the basis of which reports were produced and shared with Save the Children. However, the collaboration went cold after that possibly due to turnover in the team at Save the Children (Field notes). Meanwhile, the team continued to work on updating their model because the model for Tigray was not working as well as their previous ones and potentially highlighted a need for change in the rule sets (Interview with DS, 15/07/2021). During this time, they continued to work on reducing validation errors for these new models while updating and sharing proposed changes to the rule set. As the collaboration became active again this ongoing work could be supplemented with important contextual information and insights that came to light for example, a big population remains at the border and do not undertake border crossing as this is influenced by recommendations from armed groups. Further, that camps tend to be ethinicised and refugee's decision-making processes are influenced by ethnic composition of the camps with misinformation starting to play a significant role in route selection.

> In successive meetings Save the Children highlighted the presence of an ethnic pull factor with displaced persons moving towards camps of same ethnicity. With camps being ethnicised, refugees' route selection comes to be influenced by the ethnic composition of camps. This depends to some extent on the circulation of information among people on the move but such information circuits are difficult to pin down, understand, or capture meaningfully (Field notes - Humanitarian organisations #2; #6).

The importance of demographic information was also highlighted to tailor resources and target support. These underscored the intensely complex and multicausal nature of forced displacement. It also raised concerns around the compounding role of slow onset climate change even though forcibly displaced persons would likely never respond with climate change as a migration driver. This foregrounded the role of keeping the model based on events or frequency of events.

> For example, in Myanmar, with the flooding of the delta, houses would get washed away every 30 years while now it happens every 10 years (Field notes, example provided by Save the Children) (Humanitarian organisations #5).

Active collaboration also highlighted how team changes in humanitarian organisations hamper coordination as do issues around funding for these sort of exploratory activities along with lack of human resources in the form of a data scientist. For example, a new lead in the field office might increase responsiveness while a front-end that can be used by non-technical personnel might get more buy-in and funding allocation. When after almost a year, in April 2022, Save the Children finally had a data scientist, Ukraine had assumed centre-stage. It raised the long term question on how to integrate ABM in situations like this and it reverted back to the struggle to get adequate data even though Ukraine was a much more data rich environment compared to the conflict scenarios developed so far. In order to help Save the Children implement the code at their end, the team provided them with the code contained within Jupyter Notebooks but they ran into trouble while running the code on Windows as it ran for 48 hours without result. This highlighted an issue with deployment of the code across different operating systems since it had run quite well on Mac and Ubuntu (Field notes).

The collaboration with IOM has similarly been ad-hoc. IOM were instrumental in providing the much needed qualitative information for updating the rule set that resulted in Flee 2.0 (P13). For this instance the team shared a list of open ended questions that IOM distributed to NGOs on the ground and collected the responses they received and shared them with the team. They then reached out post a workshop on Flee held on 24/09/2021, where they got to know about the collaboration on Tigray with Save the Children, for a discussion on the granularity of IDP modelling right down to lower administrative levels. This highlighted the need for a new rule set for IDP movements as they exhibit different dynamics compared to refugees. A call with the IOM specialist teams was also arranged which helped flag the constantly changing conflict dynamics which not only made it difficult to collect data but made any analyses outdated since they would no longer be relevant in real time (Field notes). This highlighted the scope for Flee to offer simulation scenarios for operational planning when data was unavailable or inaccessible (Field notes; HiDALGO, 2021b).

5.3.3 Academic

The development of Flee often proceeded through collegial academic collaboration (both within the Department and external) beyond funded consortia to explore shared research questions or get the required expertise on a particular aspect of the problem. It was DG's joint work on FabSim with colleagues from UCL which was the basis for developing the FabFlee plugin (Groen et al., 2016). These academic collaborations formed the basis for co-developing grant proposals and building on existing work through successive grants. There were often collaborations forged between academic colleagues, consortium partners, and humanitarian organisations like the one between ECMWF for the Mali route accessibility dissertation and the one between Colombia University and Save the Children to potentially discuss the unfolding situation in Ukraine.

There was collaboration with Columbia University to integrate Flee with the US government funded World Modelers Project (https://worldmodelers.com/) (HiDALGO, 2021b) which has two components: an architecture that prepares, publishes, and updates models and an interface that helps draws insights from the combined repository of models incorporated within the architecture. The latter is to help enable analysts to draw bigger picture insights by allowing the combination and adjustment of parameters to build custom models and scenarios. The idea is that this would help combine quantitative approaches and qualitative understanding of the phenomenon to perform comparative spatio-temporal analyses to identify composite risks and plan interventions. The discussions with Columbia University revolved around how to make Flee more generic and less reliant on historical events to enable its incorporation within the Project. However, it was highlighted how introducing flexibility within the code introduces technical complexity which limits its usability by non-technical users (Field notes). This also highlighted the time consuming aspects of model building like processing open humanitarian data for input files and constructing location graphs. It helped explore complementarities in the work done for Save the Children and IOM in modelling the Tigray conflict and IDPs.

Advice on a student dissertation on geoinformation systems at the University of Utrecht in Netherlands brought the team in touch with a member of the Dutch Ministry of Defence who was the co-supervisor and a former geospatial analyst stationed in Mali (Boesjes, 2022; Boesjes et al., 2022; Field notes). The dissertation aimed to explore the role of route accessibility and paths taken by forcibly displaced persons when trying to reach safety. The collaboration helped provide detailed insight and information on weather conditions and route accessibility in Mali. Mali has alternating and dry and wet seasons; during the wet season, the valleys and routes along the river become impassable as result of which fighting shifts elsewhere. Moreover, Mali is getting drier over the years and it was suggested to look at data over the past 10 years. This has been exacerbated due to international troop deployments in the region which led to excessive groundwater extraction, leaving the region drier. Further, slope, elevation, and vegetation become important factors when it comes to locations where conflict is likely to erupt as density of vegetation and terrain conditions can make some areas impassable. This highlighted the range of different phenomenon that have to be explored in depth and factored into their models to understand the effect they have on the behaviour of the model. Even though vegetation index and moisture index might be available with the Copernicus project, this entailed detailed time taking work to construct and test them for each conflict scenario. Even though this did not translate into a functional extension for Flee, the insights gleaned from this collaboration refined understanding of terrain conditions and route accessibility for conflict progression and forced displacement.

5.4 Environment: Conditions surrounding the development of Flee

The development of Flee was conditioned by diverse factors ranging from the computational, digital, and data to domain, socio-political, and policy, legal and ethical conditions. The computational, data, and digital environment included the availability and pre-conditions of open source software, open data, and issues around data quality and accessibility in conflict zones. Domain level conditions spanned the specificities of individual conflict and location dynamics to geographical disparities in funding across humanitarian operations. Socio-political conditions like flashpoint events created both demand for predictive models while at the same time diverted attention away from existing conflicts. Demand for modelling specific scenarios continued to be sustained by an enabling policy environment for predictive computational techniques while at the same time contending with legal and ethical considerations for downstream adoption, trust, and safety.

5.4.1 Computational, digital, and data

Flee itself is an open source code and relies on open source tools, toolkits, and frameworks like FabFlee, VECMAtk (continued as SEAVEAtk), NSGA-II, MUSCLE 3; Python libraries and packages like pandas, matpotlib, MPI4Py (for parallelisation). Further, the input data for Flee comes from open data from UNHCR, ACLED, and geospatial data from Bing Maps and Open Street Maps. Keeping the software open source is useful to enable further uptake, use, and dissemination. Humanitarian organisations like UNHCR have a clear preference for open source software due to transparency and accountability requirements (Field Notes, Team Meetings #45[3]). The European Union's Open Science Policy is the standard method of working under its research and innovation funding programmes and require that beneficiaries of such grants make their research outputs openly accessibly to facilitate wider participation of citizens and stakeholders. The purpose is to ensure storage, sharing, processing, and reuse of such outputs produced through public money (European Commission, n.d.). However, it recognises significant considerations that come into play in linking open science objectives to innovation and business models in the form of intellectual property rights, licensing arrangements, interoperability and reuse of data which links to the policy, legal, and ethical environment governing the same. This tension between open accessibility and intellectual

---

[3] Specific claims have been attributed to the meetings in which they were made.

property rights came into play within the ITFLOWS project as lead developers of the large scale model, a private entity, contended against making their source code available to the other members of the consortium on the grounds of it being proprietary. However, they agreed to provide an executable version of the code which could be run to arrive at results. This led to significant dissonance within the consortium due to the ethical implications of developing a closed source software on an EU grant, the utility of linking an open source model like Flee with a closed source model, concerns about the resultant ownership of the model, and the ability to do research and disseminate scientific outputs resulting from such a process. This also raised questions on the overall sustainability of the whole endeavour in building a model that could serve a public utility purpose for the EU.

> This involved questions and concerns around who would own the overall product and be responsible for processing, updating and providing the information to relevant stakeholders beyond the life of the project (Field notes - ITFLOWS #1; #15; #16).
>
> At the time of concluding the fieldwork, the objections against closed source software, particularly by the team, was highlighted to principal investigators. The lead developers in question (for ITFLOWS) were in contact internally within their organisation to make the source code available to consortium partners under a non-disclosure agreement though the prevailing concerns still remained despite restricted access (Field notes - Team meeting #47; #48; #49; #53).

Open data can be used without legal complications (Field Notes, ITFLOWS #1) and is a common practice for developing open source software for social purposes like Flee because of ease of accessibility and onward dissemination. The humanitarian community promotes open data to enable collaborative data sharing, efficient humanitarian response planning, and innovation to ensure the responsiveness and adaptability of the community to changing needs (UNOCHA, 2023; UNHCR, 2019). The critical need for such data is often felt when demand spikes during flashpoint events like an outbreak of war or conflict or climate crisis. Conflicts in Ukraine, Tigray, Nigeria and drought and food insecurity in the Horn of Africa drove demand for data in 2022 (UNOCHA, 2023). However, data availability is strongly linked to socio-political drivers within the humanitarian context.

> The lack of availability of data results from lack of access to locations of violent conflict where risk, vulnerability, and human cost of data collection are quite high (KEI Event, 30/11/2022).
>
> Further, by the time any data would have been collected they would become quickly outdated due to the rapidly evolving nature of most conflicts and changing political scenarios. There is a time lag involved in accessing, collection, and processing the data before making it publicly available (Humanitarian organisations #4).

This is compounded by more systemic operational issues involving update of registration numbers in camps which makes a particularly important dataset like that of UNHCR difficult to validate against.

> Many of the key datasets within the sector suffer from a methodological drift with significant implications for data outputs. Updates to registration numbers on periodic review can either mean an initial undercounting or overcounting or actual changes in numbers at the camps and there is no way to know for sure (KEI Event, 30/11/2022).

Moreover, the UNHCR data portal once had API documentation relating to an older version of the portal with some of the conflicts still under the older system (Suleimenova et al., 2017b). ACLED does not provide duration of conflict but reports individual events (Field Notes, Team Meetings #50). Further, the sources of data outside of the UNHCR was in the form of reports and texts i.e. not in machine readable format (Chan et al., 2018) and could not be directly inputted into the model; they serve to inform a more tacit understanding of the conflict scenarios. Qualitative informational inputs that have general applicability across conflict were incorporated through rule set changes. Moreover, the format of data availability require significant pre-processing so that it can be fed into Flee. UNHCR data is available in JSON format and needs to be converted into .csv for it to be used in Flee. While ACLED data is available in .xls format, it contains a range of information that is outside the scope of the model (Suleimenova et al., 20717b) which prevents full automation of the workflow. Complicating the fragmented data landscape is the non-availability of data on different causal factors affecting forced displacement which have to be combined with data sources from other organisations, with different mandates like food security and weather conditions, to establish relationships that might be an operationally and qualitatively known phenomenon at the domain level.

5.4.2 Domain

Forced displacement is a multicausal phenomenon that is affected by a range of social, political, and economic factors. These include opening of borders, camps, and refugee registrations, conflicts, mineral mining and pricing, elections, slow onset climate change, demographics (families with children and elderly), religious and ethnic dimensions, travel route and accessibility, group dynamics etc. for which data is not collected or available (Chan et al., 2018; Campos et al., 2019; Suleimenova & Groen, 2020; Field notes). Moreover, there are complex relationships between food insecurity, climate change, and war with many arguments about the ways the latter influences the former. In the absence of a comprehensive multidimensional view of forced displacement, validation of computational results investigating such relationships range become untenable and need to include intuitive, operational, or secondary sources. Also as mentioned earlier, hardly any displaced person might state climate change as the cause of their decision to move. This is despite changing climate conditions in origin countries, for example, Mali has been consistently becoming drier with changing migratory patterns between North and South (ICRC, 2022; Field notes).

> In the collaboration meeting with ECMWF for the University of Utrecht project on 03/07/2021, in response to a questions by ARJ on climate refugees and migratory patterns from the North to South, the supervisor from the Dutch Ministry of Defence highlighted that it is indeed true that the region has been getting drier over the past few years exacerbated by the presence of peacekeeping missions and extraction of ground water for their upkeep (Field notes - Others #1).

While conditions of forced displacement might be difficult to model, IDP modelling faced considerable challenges because the drivers were neither fully available nor understood. IDP modelling was only undertaken through the collaboration with Save the Children due to the significant presence of IDPs in Ethiopia and qualitative information provided for urban-rural migration patterns and the role of family and friend networks within IDP migration which are difficult to determine from data alone.

Moreover, dynamics of forced migration and humanitarian operations change from country to country. Within the South Sudan conflict starvation tactics were employed by preventing access to humanitarian aid. Starvation and malnutrition were the leading cause of death for 7 million people with 4.3 million forcibly displaced out of which 2.5 million are refugees (Campos et al., 2019). This was exacerbated by famine and the local climate that alternated between dry and wet seasons with the latter causing floods with significant consequences for the population. With the Sudanese government closing the border with South Sudan, the situation worsened due to lack of free movement of residents and goods between the two countries (Suleimenova & Groen, 2020). This underscored a need to understand the effect of policy decisions such as camp or border closures, camp capacity, and forced redirection on distribution of refugee arrivals in neighbouring countries. Further, South Sudan was difficult to simulate due to lack of roads (Suleimenova & Groen, 2020). Given few roads and mountainous areas more walking routes are added and broader range of phenomenon like weather conditions were included. Mali's desert terrain and climatic conditions resulted in the use of off-road links as highlighted from the experience of the personnel from the Dutch Ministry of Defence (Boesjes et al., 2022). This was incorporated by creating a graphical representation in the form of a raster that takes into account mobility altering features of the physical environment as per NATO classifications. Progressively causal parameters were explored when such relationships could be modelled on suitable datasets like IPC in the case of food security and MSCI minerals and mining data in case of the Flare model (Campos et al., 2019; Groen et al., 2019a).

There are also structural issues in funding across different field offices with both Central African Republic and Burundi being severely underfunded refugee operations with funding shortages of 76% and 62% respectively which impacts their investment in data collection and availability (Suleimenova et al., 2017b). This in turn has an effect on the ability to construct detailed and validated models for those countries. While some of these causal relationships were explored by combining complementary data and modelling approaches in country contexts that offered the most insight there was recognition of both the limitation placed by a system as complex as forced displacement as well as the need to achieve a better approximation of it. Moreover, flashpoints events shift the focus away from ongoing operations to assume centre-stage.

5.4.3 Socio-political conditions

2015 was the year of the European refugee crisis when hundreds and thousands of people fled war and persecution to reach safety on the shores of Europe (Spindler, 2015). Despite the comparatively higher refugee flow in countries bordering conflict countries, the European crisis received more attention, where existing institutions faced significant strain (PwC, 2017). The use of treacherous routes across the Mediterranean unfolded in immeasurable tragedy and loss of lives. The image of the lifeless body of Aylan Kurdi washed up on a Turkish beach after a failed attempt to reach Greece underscored the grim tragedies in capsized boats and refrigerator truck deaths across land and sea routes or increased hostility and poor living conditions in host countries if they did manage to survive the ordeal (Spindler, 2015). This prompted a response from then UNHCR Chief Antonio Guterres to issue key guidelines for dealing with the unprecedented situation (Clayton, 2015). The guidelines emphasised the need to stabilise the current situation while working on a long term strategy for shared responsibility. This involved the need to have adequate emergency reception, assistance, and registration capacity. This is the context in which Flee was initially conceptualised. The European migrant crisis also formed the backdrop for the ITFLOWS project (ITFLOWS, 2021e) which aimed at looking at the broader context of migration to help EU institutions and stakeholders facilitate better preparedness for reception, relocation, settlement, and integration.

Like the 2015-16 migrant crisis, flashpoint events like Tigray in 2020, Afghanistan in 2021, and Ukraine in 2022 created a surge of demand for better data and actionable insights for operational planning. It also created a demand for scaling up existing models and directing attention to the modelling of these unfolding crises. With the outbreak of the war in Ukraine attention, resources, and expertise were diverted away from crisis operations elsewhere (Redfern, 2022). This redirection of efforts and resources affects international humanitarian supply chains leading to higher rates of inflation (already compounded by COVID-19) with rises in fuel and food prices increasing the cost of transporting aid and emergency medical treatment. This is particularly severe for countries in the Eastern and Horn of Africa where worsening drought coincided with the Ukraine war (Redfern, 2022). With limited resources going around, operations in these countries would be scaled down to the bare essentials and like Burundi and Central African Republic this underfunding would affect data collection and digital innovation activities. However, as the field notes suggest, rapidly scaling the model for an unfolding conflict in real time was not in tandem with the demand. This was because of the time lag in data availability and information outflow and the lack of presence of humanitarian actors in some neighbouring countries and not others. These flashpoint events drive demand not only for modelling new conflicts but also reinforce the policy push towards predictive analytics for anticipatory action in order to make the money 'go further' (Lowcock, 2019).

5.4.4 Policy, legal, and ethical

The number of people in need of humanitarian assistance soared in 2021 to an unprecedented 235 million (UNOCHA, 2020). While the quantum of humanitarian funding has increased over the years, the needs have continued to outpace them by a significant margin. COVID-19 alone caused $9.5 billion dollar requirement with the resource-needs gaps growing wider every year (UNOCHA FTS, 2021). This highlighted a need for a system level change in humanitarian action as highlighted by Mark Lowcock, the then Under-Secretary General of Humanitarian Affairs and Emergency Relief Coordinator. Lowcock underscored the need to anticipate rather than react to crises with new and emerging technologies and predictive analytics supporting this paradigm shift. The hope was that predictive analytics technologies can help make sense of the vast and complex humanitarian data to improve predictions and decision-making (Lowcock, 2019). His speech delivered at the LSE in 2019 highlighted both current and future efforts at harnessing digital innovation to assist in protracted humanitarian crises. It referred to consolidation of the predictive analytics work at the UN OCHA Centre for Humanitarian Data which has eventually come to host a comprehensive repository of predictive analytics models for humanitarian action, including Flee (https://centre.humdata.org/catalogue-for-predictive-models-in-the-humanitarian-sector/). It also signalled the increasing importance of exploring the role of digital technologies through funding, investment, and collaborations as evidenced through workshops held on the topic, outreach by humanitarian organisation to explore the technology landscape, and the significant funding outlay for a project like ITFLOWS.

However, these technologies also presented significant risks and challenges. Despite the recognised utility of predictive analytic tool there were significant concerns about the risk of harm to migrants' rights, needs, and interests (Guillen & Teodero, 2023). This speaks to broader societal concerns about the risks and harms of using predictive technologies as well as how they apply to the already fraught context of forced displacement in amplifying vulnerability and precarity. EU has one of the strongest data protection regulations in the world in the form of GDPR (General Data Protection Regulation) as well as an active policy landscape for the regulation of artificial intelligence. Therefore, technical outputs needed to be in legal harmony with regard to both the predicted phenomenon as well general data protection principles. Within the ITFLOWS project this meant that the predictive output of EMT needed to be harmonised not only in line with relevant international and EU norms on migrant and refugee legislation but with GDPR as well as with internal policies of humanitarian organisations (ITFLOWS, 2021a; 2021b; 2021c). As mentioned earlier, this becomes particularly complex due to the use of different datasets that might use different definitions of a refugee. This potentially calls the resultant output into question if the intended jurisdiction of its implementation follows a different legal and statistical interpretation that contradicts that of the dataset which provides the input data to the model. As a result, the ITFLOWS project involved ethical, societal, and data protection risk assessments which involved definitional work, identification and application of the correct terminology through a mapping exercise of definitions and datasets used (ITFLOWS, 2021b). This resulted in the adoption of the terminology ‘asylum seekers/ unrecognised refugees’ for predicting the distribution of forcibly displaced persons across camps. This was followed by a regulatory framework for the ITFLOWS project to ensure a framework for compliance with relevant laws and regulations, principled design of digital technologies, including a continuous monitoring and enforcement framework (ITFLOWS, 2021c). Ethical considerations also existed in the development of conflict prediction model for fear of misuse (HiDALGO, 2021b), instead the team has either used a conflict evolution model based on high level heuristics as in the case of Flare or forecasts based on conflict scenarios as in the case of Tigray.

5.5 Nature of the architecture, ecosystem, and environment

As mentioned earlier, the literature from digital product innovation highlights the interdependency between the architecture and the ecosystem wherein the nature of architecture organises innovation among a range of heterogeneous actors and innovation proceeds through negotiations within the ecosystem. Parallelly, complex products and contextual studies highlight the role of wider environing conditions. Therefore, theorising the nature of relationships driving the innovation trajectory involves explication of the nature of the architecture, ecosystem, and environment. Developing their thick descriptions becomes an important first step or building block for theorisation (Layer 1 of Fig.1) since the nature of the architecture organises innovation among heterogeneous actors within its proximate context wherein relationships negotiate technical and structural conditions to move towards the realisation of design choices or outcomes at each successive stage of digital product innovation. Tracing the evolution of the product from its initial stage or the first prototype model highlighted a hybrid architecture with flexible modular components that helps maximise innovation while keeping the overall model close to the central parameters and assumption in its rule sets. Further, each successive stage of development involved successive configurations of ecosystem participants depending on the design choice and the negotiation of multiple environmental drivers which enabled and constrained the scope of possible action in conjunction with hybridity of the technical architecture.

5.5.1 Hybrid architecture

The Flee code depends on a range of automation tools, frameworks, and sub-models which is subtended by a range of parameters and assumptions which comprise the Flee rule set. There have so far been 3 iterations of the rule set, each aimed at updating parameters and assumptions and improving the functional efficacy of Flee, with the aim of making it more robust and relevant to the complex social reality it attempts to capture. Changes to the Flee code requires significant specialised knowledge about the code and the overall workflow while the output is sensitive to the input data and parameters. However, the code itself is often dependent on more generic modular components like FabFlee for automating work processes, Flare sub-model for forecasting based on conflict evolution, as well as MUSCLE 3 for coupling multiscale simulations. FabFlee is based on the FabSim 3 architecture which contains seven modules, six of which are easily modifiable by users where modules are extended by defining new Python functions with domain specific scripts and other resources. This underpins FabSim's multidisciplinary adaptability and its iteration in the form of FabFlee (Groen et al., 2016).

Table 2: Hybrid Architecture Generic extensible software components	Phases	Integral features	Phases
FabSim	P3	ABM framework	All
FabFlee	P7, P8, P9, P13, P14	SDA	P2, P3, P6, P7, P8, P9, P10, P13, P14, P15
VECMAtk (continued as SEAVEAtk)	P13	Flee rule set	All
MUSCLE 3	P10		
NSGA II	P14		
Python libraries like and packages like pandas, matpotlib, MPI4Py (for parallelisation)	P1, P2, P5, P11		

Parallelisation of the code in multiscale modelling enables the mix-and-match of multiple at-scale models which also underpins the MUSCLE coupling environment (Groen et al., 2019b). The Flare sub-model too is modular in nature as it allows for the incorporation of algorithms from external parties (Groen et al., 2019a). However, these multiple generic extensible software components which form a part of the Flee's computational environment does not change its integral composition and its output remains highly sensitive to input parameters. Table 2 (Hybrid Architecture) shows FabFlee has been a key component for extending the functionality of the product in the form of coupling, ensemble simulations, rule set updates, and simulation-optimisation approach. It is underpinned by the modifiable FabSim architecture which allows the code to easily add and adapt additional functions. Similarly, components like VECMAtk, NSGA II, and python libraries have been instrumental in extending product functionalities in the form of SDSD approach, multiscale migration modelling, simulation-optimisation approach, and in developing the initial prototypes respectively. However, the Flee code is based on the ABM framework and contains a number of interdependent parameters and assumptions within its rule sets that determine the behaviour of the agents across time and space within a given set of conditions. This means any functional extension brought about by generic digital components would need to either operate within the constraints of the integral components or require changes to the integral components themselves.

One of the ways in which Flee has attempted to move towards more general use has been through the formalisation of the approach through SDA. The formalised approach aims to promote uptake and use of the model by guiding its application across contexts using the framework and software package offered by Flee. Flee code itself builds on multiple inheritances from previous iterations. The original model after being formalised through the

SDA underwent multiple iterations and extensions wherein the Flee model itself was coupled with sub-models such as the food security and conflict evolution sub-model (P7 and P8). The rule set too has undergone changes to incorporate information and insights provided to bring the model in closer approximation with reality. These changes became important when the functional changes required could not be handled using generic components alone such as in the development of SDSD approach and rule set update (P13) and integrated IDP modelling (P15). On other occasions, the SDA has been extended to accommodate hybrid simulation modelling and counterfactual scenarios like in P8 and P9.

The Flee code involves combination and translation of input data and parameters into potentially actionable outputs for decision-making through computational code, functional commands, and scripts. These, in turn, work within the constraints and parameters determined by the rule sets and implemented through the SDA within an ABM framework. These technical interdependencies within the architecture allows the code to harness, execute, and produce results across different parameters and conditions. This combination of generic dependencies along with integral nature of Flee exhibits a form of hybrid architecture.

A hybrid architecture has implications for how the ecosystem is organised and structured around it. It leverages the generativity and modularity from the computational tools and frameworks within its own domain of application but transmutes them into tight couplings within the integral components of the Flee code. It helps make the integral architecture more flexible and amenable to experimentation by enabling recombinability of computational resources to aid in innovation. The integral architecture harnesses this modularity to produce, automate, or bring into effect the changes implemented within the code.

5.5.2 Successive configurations of ecosystem participants

The hybrid architecture requires successive specialised multi-actor collaborations drawn from the ecosystem to resolve particular design issues. Table 3 (Actor configurations) shows how different multi-actor coalitions were formed at different phases to arrive at associated outcomes.

Table 3: Actor configurations

Successive multi-actor configurations	Phases
Single developer	P1, P5
Team	P9
Team, Extended academic network	P2, P3, P4, P6
Team, Extended academic network, Consortia	P7, P8, P14
Team and HiDALGO	P10, P11, P12
Team, Consortia, Humanitarian Organisations	P13, P15
Team, Columbia University / University of Utrecht	Not tied to a specific outcome.

As the table shows, model development proceeded through successive configurations of different actors in the ecosystem and almost always involved the team along with other actors to leverage multi-specialised knowledge. A single developer or solely the team was involved in developing prototypes like the initial model (P1) or parallelisation and multiscale coupling (P5). P9 or exploring the effect of policy decisions through ensemble simulations were a part of DS' PhD thesis which was later refined. In the initial phases model development was solely driven by team members with parameters and assumptions based on intuition with computational and socio-political drivers at play due to limited stakeholder connections. Moving forward from the initial stages, advancements were made through extended academic network and collegial collaborations. It involved academic collaborations within the same Department, serendipitous network opportunities, introductions by former colleagues, and conferences and workshop participations like the Science Hackathon Geneva hosted in 2017 by CERN and Collaborations Workshop 2017 hosted by the Software Sustainability Institute which contributed to FabFlee automation phase (P2, P3, P4, P6). The ensuing collaborations despite loose ties helped established the foundational basis for Flee that attracted the eventual funding. Such loose but substantive ties were also observed in the case of collaboration with humanitarian organisations. Collaborations with humanitarian organisations could be ad-hoc with long periods of gaps followed by periods of intense collaboration. These collaborations also elicited significant contributions towards rule set changes in terms of parameters and assumption that determine how the code operates. It also contributed to tacit knowledge about granular operational conditions, thereby potentially opening up future areas of explorations. It also helped expand the scope of Flee to include forecasting of a current conflict and modelling IDPs while also contributing a major development in the form of rule set change resulting in Flee 2.0 (P13, P15).

The collaboration with Columbia University helped highlight important decisions to be made with regard to Flee's architecture with regard to flexibility, usability, and integration with a suite of models for composite analysis. The collaboration with the University of Utrecht on the student dissertation on route accessibility in Mali elicited critical contextual information that shaped the understanding of route construction and the relationship between seasonal changes, conflict dynamics, and route accessibility. These two collaborations, despite not culminating in substantive outcomes, informed the innovation process. These examples highlight that despite the strength of the ties, the nature of their contribution substantively shaped the development process.

While these loose connections facilitated important developments and advancements, these would not have been possible without the strong ties engendered by funding consortia. HiDALGO and VECMA continued to shape the development process through financial and human resources as well as their project mandates. These strong ties were shaped by project objectives that mandated particular kinds of scientific explorations. Strategic manoeuvring in aligning project mandates and research directions shaped by the tacit understanding gained from the above collaborations translated into the evolution of Flee across different phases such as P7, P8, P10, P11, P12, and P14 by including additional causal conditions. This was particularly important because while the loose connections shaped the requirements and desired direction of taking Flee towards practical use, the strong ties forged by funded projects helped actualise some of the intended directions. This was made possible not just through passive financial support for maintaining the core team but also through the required expertise and advanced resources like supercomputers, e.g. for scaling parallelised version of Flee (P11). This included collaborations with ECMWF and KNOW Centre to explore relationship between weather conditions, conflict, and displacement and improve construction of location graphs through automation (P10, P12). Within the VECMA project it helped in integrating sensitivity analysis within SDA to mathematically evaluate the strength of causal relationships (P13). Both helped test the performance of the code on supercomputers and test their reliability through detailed and multiple runs. The ITFLOWS project provided support for increasing the model portfolio while providing interdisciplinary support to improve legitimacy of the outputs of the model for decision-making (P14). Thus, the ecosystem provided the conditions to form specialised multi-actor alliances through co-dependent relationships between strong and loose connections the team had with different stakeholders which would translate to architectural development on their suitable alignment.

5.5.3 Multiple environmental drivers

Resources and relationships unfold within the context of different environmental drivers identified in Table 4 (Environmental drivers) below. These drivers work in tandem with the project mandates and ecosystem conditions to determine the combinations of different resources during given phases. While distinct environmental drivers have been identified, they do not work in isolation and the combined inter-locking effect of different drivers shape the resource limitations and trade-offs within the development process.

Table 4: Environmental Drivers **Multiple environmental drivers**	
Computational, data, and digital	Open source tools, toolkits and frameworks; open data; open science policy
Domain	Multicausal nature of forced displacement; differing conditions for refugees and IDPs; country specificities; non-availability of granular data; seasonal variations in route accessibility; funding inequalities across humanitarian operations in different geographies
Socio-political	EU refugee crisis; flashpoint events – Tigray, Afghanistan, Ukraine
Policy, legal, and ethical	Anticipatory action policy; data protection; ethical concerns, legal harmonisation

Socio-political drivers like the migration crises can create funding priorities which both motivate the development of Flee and create institutional funding conditions for their eventual support. While computational, digital, and data drivers relating to data quality place resource limitations, flashpoint events by creating a demand surge lead to greater informational availability through active collaborations. Further, a policy environment predicated on a computational and predictive future of humanitarian action geared towards optimisation of resources provide the required impetus and involvement of greater number of potential users becoming invested in its development towards operational implementation.

Multiple environmental drivers stand in complex relationship with the architecture and ecosystem. While some drivers like domain and socio-political provide a significant push towards rapid simulation development, other aspects work to limit the extent to which complex causal relationships find formalisation through computational models. This constant negotiation foregrounds both the nature of active ties withing the ecosystem and the aspects of the architecture such relationships can contribute to. While modular aspects of a hybrid architecture admits additional details to be incorporated within the model, the integrality of computational algorithms and mathematics structure the way causal relationships can lead to reliable validated outputs. While weak ties can translate to fundamental changes, stronger ties help in providing the output validity by translating qualitative information into a quantifiable score of reliability. The following chapter unpacks the nature of this complex multilateral relationship through the adapted choice framework to understand how mediating conditions identified below shape, resolve, and negotiate the attendant tensions.

5.6 Choice components

Application of the adapted choice framework involves unpacking successive outcomes or design choices by identifying mediating conditions in the form of the resource portfolio, interpersonal relationships and structural conditions which help create the basis for understanding the interrelationships between agency and structure in relation to technology (Layer 2 of Fig. 1). Within the adapted choice framework, outcomes become the starting point for socio-technical enquiry to unpack the role of different resources, actor configurations, and environmental conditions that jointly shape both a given outcome and the process as a whole within the opportunities and constraints of the technical architecture and how it organises the innovation process.

5.6.1 Resource portfolio

Application of the adapted choice framework involves identification of different resource categories that become the basis of exercising agency in relation to structure and technology. The following categories of resources have been identified as contributing to its development (also see Appendices 1 and 2). Identification of these resource categories help highlight their role within the multilateral interdependencies that shape the innovation trajectory. It also helps understand how different combinations of these resources help bring about particular development since their availability in specific combinations are a function of the interdependent nature of Flee's development. However, the development process not just used different resources but also resulted in the creation of new resources that went on to being used in successive development processes.

(a) Computational resources

Computational resources were resources that digitally or computationally supported the development of Flee. These include software suites, toolkits, or frameworks as well as hardware like supercomputers. During the course of its development, Flee used a range of computational resources that included:

- Python libraries like pandas, matpotlib, numpy, scipy for data analysis, visualisation, and for incorporating advanced mathematical functions and relationships (P1, P2, P11)
- FabSim 3 to develop the FabFlee plugin for a range of functions spanning automation, sensitivity analysis, coupling, and parallelisation on supercomputers (P3, P7, P8, P9, P13, P14)
- MPI4Py (Message Passing Interface for Python) for implementing the parallelised version of Flee. (P5, P11)
- MUSCLE 3 for multiscale migration coupling (P10)
- VECMAtk, a suite of software tools including EasyVVUQ, FabSim 3, and QCGPilotJob for integrated sensitivity analysis and remote execution on supercomputers for large and detailed runs (P13)
- Supercomputers like Eagle, Hawk, and Vulcan were used to run large and detailed simulations as well test performance and efficiency of the code (P8, 10, P11, P13)

While Python libraries and software were available open source, scare computational resources like supercomputers were provided through funded consortia. This included Eagle and Altair hosted by PSNC, a consortium member in both HiDALGO and VECMA and Vulcan and Hazelhen hosted by HLRS, a consortium member in HiDALGO. Rarely simulations were also run on supercomputers that were not a part of the consortium. For example, the HPE Apollo supercomputer was used by the KNOW Centre to test the performance of the route pruning algorithm (P12). Computational resources like FabFlee that were developed as a part of the initial development became instrumental in driving successive stages of its evolution particularly for P7, P8, P9, P13, and P14.

(b) Data resources

Flee primarily used data resources that were derived from open source datasets provided by UNHCR and ACLED as well as geospatial data provided by Bing Maps and Open Street Maps which were key data resources across it phases. It also used data from population databases like CityPop and national census databases where available. Though in conflict countries, the latter tends to be unreliable or incomplete. While the UNHCR data helped in validating the results of the simulation to calculate error rates, ACLED data provided the conflict locations in which agents were spawned and helped provide the starting point from which agents traverse the location graphs constructed with geospatial data. In coupling instances, additional datasets were used like IPC for food security (P7), MSCI World Metal and Mining Index for the Flare conflict evolution model (P9), and ECMWF weather data for multiscale modelling (P10). While UNHCR and ACLED data are available open source, ACLED requires a unique download key for the user, which can be obtained by registering for free with the ACLED Access Portal. However, since accessing ECMWF data depended on a range of technical interdependencies and software packages, they were mediated by the consortium partner who also advised on the type of data points and dataset available given a particular modelling criteria. However, data resources like that of UNHCR suffered from significant limitations due to revisions and updates that made them difficult to validate against. Further, as was highlighted during the Mali route accessibility project, Open Street Maps does not capture all routes effectively since they are built by volunteers using open source satellite data and as a result are not very granular. Important routes that are in use and their changing seasonal usage patterns are not adequately captured by Open Street Maps.

(c) Knowledge resources

Developing Flee involves range of software development and mathematical approaches to formulate relationships and create frameworks for future simulation construction. Existing ABM frameworks and simplified simulation development processes were used to develop the SDA for quicker simulation construction and has been foundational for the future development of Flee across its successive phases (Table 2: Hybrid Architecture). Along with this, mathematical approaches were incorporated to explore a range of different relationships between variables as well as in calculating error rates as a part of validation processes. This ranged from simple mathematical correlations to advanced forms of sensitivity analysis in the form of Sobol indices and stochastic colocation. The knowledge resources created during the development process helped guide ongoing development and explore new relationships, for example, the multiscale modelling approach (P10) helped in coupling additional factors while the hybrid simulation approach used in Flare highlighted how two different functional models could be coupled together (P8). Advanced mathematical knowledge resources resulted from the pooling of expertise through funded consortia, particularly the VECMA project. Some knowledge resources which had not found purchase in previous iterations were eventually transmuted to new knowledge resources, for example, the FLP approach used in gamification (Estrada et al., 2017) (P4) was eventually used to develop the simulation-optimisation approach (P14). In conjunction with computational and data resources, knowledge resources aimed to capture relationships, quantify approximations and divergences from reality, or further onward development. Knowledge resources were augmented by additional human resources by pooling diverse expertise and constrained by the computational and data environment which in turn limited the aspects of reality that could be modelled to those that could be represented though mathematical relationships.

(d) Information resources

Information resources and insights provided by humanitarian organisations helped make significant updates to Flee rule sets to inform understanding of displacement dynamics. Information resources helped overcome the limitations of data deficits and mathematical conditions in shaping how Flee works in fundamental ways through rule set updates. Information and insights from NGOs on the ground facilitated by IOM helped move towards Flee 2.0 through integrated sensitivity analysis and updating parameters to account for walking, movement speed, and rest stops (P13). The collaboration with Save the Children helped identify further areas of exploration out of which support for demographic attributes were in the process of implementation in the first iteration of Flee 3.0. Information resources take a long time to translate into rule set changes as they have to undergo multiple experimentation and testing including significant technical development to be formalised as rule set updates. While information resources help overcome other resource deficits, their technical codification is dependent on the extent to which certain attributes can either be represented or validated by data. For example, while the role of ethnic preferences in route selection are operationally known, it might be difficult to incorporate that aspect in the absence of corresponding data. Even if it was synthetically constructed with placeholders it would fail to pass the test of mathematical VVUQ approaches to assess the reliability of the model. Apart from information resources provided by humanitarian organisations, the team relied extensively on secondary resources like news reports and publications to develop an in-depth understanding of conflict conditions and dynamics.

(e) Financial resources

Funded consortia provide the financial resources to hire team members and expand the core team. However, the initial phase of the model development including generalisation, and automation were done in the absence of funded projects. This highlighted the importance of a working prototype to attract future funding. Even though funded project came with their own project mandates of advancing given scientific criteria, they also provided the liberty to explore lines of enquiry through cross-collaboration across consortia as well as developing use case applications through collaborations with humanitarian organisations. These cross-collaborations helped leverage collective financial resources in a way that shaped the development of Flee to be more than the sum of its individual parts. It not only helped expand the core team but also helped pool collective complementary expertise within a given consortia which could be leveraged through collaborative development. They also facilitated the access to scarce computational resources like supercomputers.

(f) Human resources

The human resource requirements for different phases were either driven by academic collaborations or through funded projects that provided support for team expansion or collaboration among consortia partners. Human resources helped bring together the required knowledge, know-how and expertise to translate other resources into tangible outputs. They helped expand the scope of development as was seen with the joining of new core team members. They also helped explore new lines of enquiry through collaboration with different stakeholders like consortium partners, humanitarian organisations, and academics. While financial resources helped provide sustainability for maintaining the core team and ensuring continued development, human resources are not always contingent on funding. This is particularly evident in unfunded collaborations between humanitarian organisations and academics. Though the financial resources might not be the underlying motivation in these cases, they do provide the enabling conditions to carry them out.

Table 5: Resource Portfolio

Resources	Agency	Limitations
Computational	Functional extension, scalability, evaluating efficiency	Hybrid architecture, resources like supercomputers are scarce and expensive
Data	Model construction	Revision of open datasets, file formats, incomplete portal migration
Knowledge	Simulation construction, incorporating multiple causal conditions, evaluating nature and strength of relationships	Cannot encapsulate full extent of reality, computationally expensive and time-consuming
Information	Parameters and assumptions in rule sets	First-hand information can be ad-hoc
Financial	Team expansion, contributes to other resource categories	Project mandates
Human	Multi-specialisation	Lack of reciprocal interest, fixed and dynamic relationships, ad-hoc

Resource portfolio represents the conditions whereby agency is exercised and represents a ‘structure of means’ whereby certain actions are enabled over others. Since resources co-occur unevenly, agency is not unfettered and is shaped by resource limitations. Table 5 (Resource portfolio) highlights how agency is circumscribed by constraints arising from such conditions. While generic software components enable functional extension they are limited by the hybrid nature of the digital architecture which needs changes to be responsive to its integral aspects. Similarly, supercomputers can enable scalability and efficiency evaluation but are scarce and expensive with access limitations. Open data can enable quick model construction even in the absence of financial resources but are plagued by methodological drift, incompatible file formats, and incomplete portal migration. Knowledge resources enable simulation construction, coupling the model with additional causal conditions, evaluating the strength of relationships through error rates, sensitivity analyses, correlations among others. However, they cannot accommodate the full extent of reality and incorporation of additional parameters can make the model computationally expensive and time consuming to run. Information resources translate to parameters and assumptions underpinning rule sets. In the absence of primary first-hand information from humanitarian organisations which can be ad-hoc, secondary sources and intuition play a significant role in shaping initial understanding of the phenomenon to be modelled. Financial resources enabled expansion and sustenance of the team while contributing to other resource categories like knowledge and human. However, they came with pre-defined project mandates within which Flee had to operate (Section 5.3.1 highlights the project objectives of different funding consortia). Different human resources enabled the bringing together of multiple specialisations required for the development of Flee but could be stymied by the lack of mutual interest as in the case of P4 or be dependent on ad hoc relationships like that with humanitarian organisation for critical information resources.

5.6.2 Interpersonal relationships

Distributed development enabled by digitality facilitates the integration of contributions from multiple heterogeneous actors. This leads to the convergence of multiple stakeholders complemented by interpersonal and professional relationships along with high user involvement from humanitarian organisations which is characteristic of complex product development. As highlighted in the adapted choice framework, given the architecture organises innovation among heterogeneous actors within the ecosystem, the interpersonal conditions within the proximate context become the mode whereby the structure of means created by the resource portfolio, technical, and environmental conditions are negotiated. Interpersonal relationships spanned academia, fixed relationships within funded project teams, dynamic relationships with humanitarian organisations, and serendipitous collaboration within the wider network to strengthen and draw synergies with individual efforts as highlighted in Table 3 (Actor configurations). Interpersonal relationships maximised with the evolution of product functionalities as is observed in successive phases of the product's development. These collaborations also brought wider resource availability within the ambit to expand collective agency to resolve design issues and work towards certain design choices.

This is illustrated in Table 6 (Interpersonal relationships) below through an illustrative comparison of three different phases. P2 which is an initial part of the development process involved open data, available mathematical knowledge, driven by intuition and academic collaborations within the same Department. These collegial relationships within academic networks helped develop the foundational aspects of the product on the basis of which further collaboration and funded projects could be solicited. Consequently, P10 and P13 towards the middle and later part of the development highlight how a funded consortia like HiDALGO enabled collaboration with consortium partners like ECMWF and PSNC. Latter part of the development from P10 onwards also saw collaboration with KNOW Centre (P12; output of P12 in the form of route pruning algorithm was later used in P14), PSNC (P10, P11, P13), HLRS (P11), Save the Children (P15) , and National Technical University of Athens (P11) who introduced additional data, financial, information, knowledge, computational and human resources to drive Flee's development through its project mandates of developing advanced computational techniques and simulation frameworks or through reciprocal utility as in the case of Save the Children. Similarly in P13, expertise from the VECMA project helped in developing the SDSD approach while IOM provided information resources to update rule sets that resulted in release of Flee 2.0.

Table 6: Interpersonal relationships

Resource portfolio:	**P2**	**P10**	**P13**
Computational	Flee	MUSCLE 3, Eagle supercomputer	VECMAtk (EasyVVUQ, QCG-PilotJob), FabFlee, Eagle supercomputer
Data	UNHCR, ACLED, Bing Maps	ECMWF Climate Data Store, GloFAS Data (Copernicus project), UNHCR, ACLED, Geospatial	UNHCR, ACLED, Geospatial
Knowledge	Simplified simulation development process (Heath et al., 2019), Average relative difference for validation, Mean Absolute Scaled Error (MASE)	SDA, Multiscale simulation approach, File I/O, Acyclic (one-way coupling), hybrid simulation approach	Sensitivity analysis (Stochastic collocation and Sobol sensitivity indices), SDA
Information	Intuition	Secondary	IOM and NGOs on the field
Financial	-	HiDALGO	VECMA and HiDALGO
Human	Team (DS joining as PhD student), Academic	Team (AJ joined), HiDALGO, Expertise on MUSCLE3 (Lourens Veen from the Dutch e-Science Centre)	Team and VECMA
Interpersonal relationships	Team, Academic (BUL)	Team, HiDALGO	Team, VECMA, HiDALGO, IOM

Fixed relationships within project teams were bound within project mandates and requirements. However, efforts were made to drive complementarity between fixed and dynamic relationships in order to move the model closer to requirements of end users like humanitarian organisations. This enabled the team to maximise and balance resources and relationships from funded consortia with requirements specifications and information resources from humanitarian organisations who were the intended final users of the model. While resources are inputs that enable the exercise of agency, interpersonal relationships also ensure wider resource availability which works towards expanding collective agency and thereby the nature of choices that can be materialised or actualised. With resources co-occurring unevenly, interpersonal relationships can affect the scale of the resource portfolio that come to define the agency component in relation to structure and technology. Unpacking the nature of interpersonal relationships within an ecosystem helps understand how they are shaped by wider technological systems and architectures and understanding the nature of involvement of different actors and the critical roles they play within a web of functional relationships (Yoo et al., 2010; Alaimo et al., 2020; Kallinikos et al., 2013).

5.6.3 Structural conditions

Structural conditions that enable and constrain agency span formal and informal laws, regulations, customs, institutions, and policies (Alsop & Heinsohn, 2005). Structural conditions relate in complex ways to the resource based input that shapes agency and action. Agency and the ability to achieve desired goals is conditioned by structural factors in the form of social, economic, and political conditions. Drawing on the adapted choice framework, this involves identifying the environing conditions around the phenomenon. Multiple ecosystem participants coming from a diverse range of contexts combined with high user involvement, derive purpose and meaning from the contexts in which they are embedded, and lead to multiple recursive inheritances from different contexts (Winter et al., 2014). They determine the feasibility and appropriateness of actions which are then circumscribed by the opportunities and limitations posed by the technical architecture.

Table 7: Structural Conditions

Environmental Drivers	**Opportunities**	**Constraints**
Computational, data, and digital	Model development even with limited financial resources through open source software and open data, scalability on supercomputers	Can come in conflict with business processes like intellectual property rights, lack of high quality data; scarce nature of resources like supercomputers
Domain	Modelling for contextual specificities	Lack of investment in data collection, methodological drift
Socio-political	Demand conditions for modelling	Flashpoint events divert attention from existing conflicts
Policy, legal, and ethical	Policy priorities and funding availability, data protection, legal harmonisation	Open data from multiplicity of sources can have different definitions of legal categories like refugees

The environmental drivers discussed in Table 7 (Structural Conditions) demonstrate how particular drivers enable or constrain the exercise of agency based on resource inputs towards particular outcomes. These stand in complex relationships with the resource portfolio as their availability is shaped by structural conditions. The environmental conditions highlighted above help identify the structural conditions wherein for example, demand conditions are created that lead to policy and funding priorities (e.g. the migrant crisis precipitating in anticipatory migration management in projects like ITFLOWS), development of dynamic relationships (e.g. conflict in Tigray leading to more focused collaborations with Save the Children), as well as the quality of resources available (e.g. information resources becoming readily available within conditions of focused collaboration leading to rule set updates) for the modelling process which have knock-on effects on the design choices or outcomes that determine the functional extension of the product. However, they place constraints on model development by diverting attention from existing conflicts, methodological drifts in the datasets that affect data quality, and differing legal definitions of terminologies used in constructing datasets by different organisations. As a result, just as they determine the feasibility of the modelling process such as through the availability of open source components, they also constrain it on account of lack of quality data that can help validate the results. These situations also highlights the conditions whereby gaps in certain resources are made up through others as a result of the mutually shaping relationship between technical and social elements within the proximate and distant context of the phenomenon, for example, data gaps being overcome by information resources gleaned from the humanitarian organisations which are then incorporated within parameters and assumptions that determine how the model operates as discussed in section 5.6.1(d) of this chapter. These structural conditions influence the nature of agency and shape the dimensions of what is achievable and realisable. The combination of structural conditions and agency jointly determine the creation of new resources like FabFlee and SDA (Appendix 2) which go on to become instrumental in determining the nature of the architecture and anchoring further innovation. These shape the feasibility of successive multi-actor configurations to take place and negotiate and navigate the structure of means provided by the resource portfolio.

5.7 Summarising key observations

The above discussion highlights the multi-technology, multi-actor socio-technical undertaking within complex digital product innovation and the nature of socio-technical interdependencies horizontally across the ecosystem and vertically across different environmental conditions. Mutually shaping role of the technical and the social within computational design, organisational arrangements, and management and negotiation of environmental conditions determine the innovation trajectory of the complex digital product under study. While the process approach identifies distinct phases of Flee’s evolution, the reality is often that of a circular process that is mutually dependent, mutually informing, and mutually reinforcing. This is highlighted in how resources created in some phases are incorporated in subsequent phases for model development and how interdependency exists among different resource categories (Appendix 2). Certain phases within the process have been foundational to Flee's development like the initial prototype, generalisation, and automation. These have anchored subsequent development and have facilitated and enabled Flee to be positioned for funded scientific projects that could help further its evolution in terms of achieving practical relevance and accuracy. The scope for generalisation offered by SDA helped establish the process for constructing and implementing models across different conflict scenarios while retaining the scope for refinement. FabFlee has been instrumental not just in the automation of workflows but in implementing counterfactual scenarios in ensemble modelling, coupling, sensitivity analyses, and the simulation-optimisation approach.

The development of Flee was anchored by the human resources that could be sustained through financial support provided by funded projects or the availability of whose expertise were facilitated through the same in the form of consortia members. It has also depended on the open source availability of a number of computational resources like FabSim3, MUSCLE 3, and VECMAtk that have facilitated its initial development and continued functional explorations. More advanced exploration was enabled by scare computational resources like supercomputers provided through funded consortia. Flee also relies on mathematical knowledge resources to formalise causal relationships and render them computationally executable. This underscores the limitations which structure the extent to which causal conditions within the extant reality can be abstracted within the model. Such constraints are also limited by the nature and quality of the open data resources that are available. This foregrounds the tension between the constraints on the modelling process and their scope of expansion on the basis of the informational resources provided by humanitarian organisations who are at once the potential end users with requirement specifications as well as stakeholders in the model development process.

This highlights the trade-offs that shape the model development process to make incremental progress towards making the model relevant and accurate for humanitarian operations. For example, the initial prototype models were always optimised for simplicity to quickly develop a proof-of-concept which was refined in due course. The first paralellised version of

Flee prioritised simplicity over scalability to investigate the extent to which the code could be scaled while retaining a simple code base. Simplicity also helped in reducing computational costs. This version was further refined through performance evaluation to identify aspects that needed further improvement. However, oversimplification to minimise the time taken by the simulation to run could make the code erratic and increase the variance. For example, normally the simulations take 30-60 minutes to run, the time could be reduced by dividing the agents which would hamper the stability of the code (Field notes, Humanitarian organisations #1). Often, despite comparatively high error rates in certain executions, the models are not refined to optimise them to avoid overfitting to a particular context and minimising its relevancy to others. This is to ensure generalisation treads a fine line between accuracy and flexibility. Similarly, while the SDA is optimised for easy adaptability across different conflicts, policy choices, and assumptions this should not limit accuracy which is why the different parameters are given continued attention to ensure they yield reliable results while maintaining generalisability. However, increasing the accuracy by adding additional details to the model has an impact on computational cost and efficiency. Increasing the flexibility of the code forecloses participation of non-technical users, requiring technical personnel to mediate model use. This compounds its ultimate objective of last mile usability by under-resourced humanitarian organisations, particularly field offices. These highlight the interdependent negotiation of these conditions across successive stage of development, the implications of which for the innovation process in unpacked in the next chapter.

6. ANALYSIS

Chapter summary: *The previous chapter traced the development process of Flee ABM. It mapped actors within the ecosystem and environmental drivers across key phases of the development of the technical architecture. This helped identify the nature of the architecture, ecosystem, and environmental conditions along with different resource categories, interpersonal relationships, and structural conditions within the innovation process. Based on these two building blocks, this chapter analyses the findings and answers research questions aimed at developing a contextualised understanding of digital product innovation. Taking digital product innovation as a process, the mutually shaping role of technical and contextual interdependencies is understood through operational research questions into its constituent aspects: coevolution of the architecture and its proximate context of the ecosystem and the interrelationship between the architecture, proximate context, and the different environmental drivers in its distant context. Delineating the environing conditions around the phenomenon of digital product innovation as proximate and distant context helps highlight the nature of the inter-locking interdependency between social and technical elements. The analysis shows that the combination of architecture, ecosystem, and environment structures and bounds generative potential within a deliberative – evolutionary process of complex digital product innovation where the ecosystem mediates attendant tensions, opportunities and constraints presented by architecture and wider environment through complementary resource – relationship configurations.*

6.1 Analysing using the adapted choice framework

As summarised in Fig. 1 contextualising digital product innovation through the adapted choice framework contains four key aspects that build on each other. Section 5.5 and 5.6 of Chapter 5 explicated the first two layers in the form thick descriptions and mediating conditions that serve as building blocks for analysis. Identifying choices linked to successive outcomes or phases in the development of Flee helped unpack the nature of the architecture, ecosystem, and environment. It helped disaggregate elements such as resource portfolio, interpersonal relationships, and structural conditions that lead to the materialisation of given choices and outcomes. Building on Layers 1 and 2 of Fig. 1, this section aims to deploy the components of analysis (Layers 3 and 4 of Fig. 1) to understand and explain the transmutation of these conditions into outcomes within and across phases. Developing a contextualised socio-technical theory of digital product innovation involves identifying and understanding the interdependency between social and technical elements of actors, task, structure, and technology (Leavitt, 1964; McLeod & Doolin, 2012; Sawyer et al., 2003; Doherty & King, 2005). It involves looking at wider conditions of possibility beyond individual economic self-interest, individual preference, and technical reasoning (Sen, 1999). However, as highlighted in the Chapter 3 (Theoretical framework), contextual research within IS has to grapple with defining what counts as the appropriate context (Avgerou, 2019). Defining the appropriate context involves negotiating between scale and detail because as the layers of context beyond the immediate setting are brought into analytical purview, the level of detail starts to recede and vice versa. This complicates efforts to explain reciprocal trajectories within digital product innovation given the open-ended nature of digitality, multilateral inheritances within its development ecosystem, and diversity of environmental conditions. As mentioned in the Chapter 3, this highlights the need for a layered – relational approach of understanding the phenomenon of digital product innovation, particularly for complex digital products like ABM. This is because the distributed development enabled by the hybrid digitality facilitates the convergence of multiple stakeholders complemented by interpersonal and professional relationships while at the same time being responsive to and conditioned by different environmental drivers as identified in the last chapter (Chapter 5).

The Findings described the nature of the architecture, ecosystem, and range of environment drivers that shape the problem space, the demand condition, and opportunities and limitations of the resource portfolio. It demonstrated how the hybrid character of the architecture provided both the space for further innovation as a result of successive digital inheritances (Faulkner & Runde, 2019) while the conditions of integrality required a more deliberative approach for updating system-wide assumptions. It also highlighted the enduring tension of abstracting a complex social reality through computational and mathematical links and relationships, requiring specialised multi-actor alliances drawn from its ecosystem to propel innovation by resolving design issues in successive phases. Therefore, while the more generic components expanded the scope for participation and experimentation, this potential needed to be carefully organised to accommodate for the sensitivity of the integral nature of the overall product. The ecosystem which comprised of funded scientific consortia and looser ties with the wider academic community and humanitarian organisations in the form of academic relationships, professional networks, specialised expertise, and ad-hoc collaborations functioned as the immediate or proximate context for the development of Flee.

The environmental conditions which included multiple drivers spanning computational, digital, and data; domain specific conditions; socio-political conditions; and policy, legal, and ethical considerations formed the distant context. While the drawing of these two contextual boundaries helps define the scale of contextualisation, the findings showed how multiple planes of influence and diversity of conditions span these layers of context in terms of the cascading effects of opportunities and constraints presented by the environment and resource limitations shaping the resolutions of attendant tensions through trade-offs. The delineation of these two contextual boundaries helps in understanding the unfolding of digital product innovation within a given scale and range of detail, particularly how conditions within and across contexts become important for developing holistic explanations for the phenomenon of digital product innovation.

Choice components of resources, relationships, and structural conditions operate across the two layers of proximate and distant context in relation to the architecture. Their transmutation into given choices and outcomes depend on the scope of contextual conditions, partial resolutions, and non-economistic considerations (Layer 3 of Fig. 1):

Scope of contextual conditions: Scope of contextual conditions refers to the variety of contextual conditions present within a given phase for a particular choice and outcome to materialise. As highlighted by Tables 5 (Resource portfolio), 6 (Interpersonal relationships), and 7 (Structural conditions), successive phases involved different resource combinations, successive multi-actor groupings, and multiple structural conditions at play. However, different phases and associated outcomes were realised through optimal combinations of these conditions. As illustrated in Table 3 (Actor configurations), prototype versions of the model like P1 and P5 were developed solely by DG to serve as simple working models or proof-of-concept. Each of these versions were later built upon and expanded in the course of the development process for Flee. Table 6 (Interpersonal relationships) highlighted expanding interpersonal relationships with product evolution that led to wider resource availability thereby widening collective agency. Table 8 below compares P5 with P10 and P11 to illustrate the scope of contextual conditions for successive stages.

Table 8: Scope of contextual conditions	**P5**	**P10**	**P11**
Resource portfolio:			
Computational	MPI4Py	MUSCLE 3, Eagle supercomputer	Hawk, Eagle, Vulcan (HiDALGO), pflee (parallel algorithm)
Data	-	ECMWF Climate Data Store, GloFAS Data (Copernicus project), UNHCR, ACLED, Geospatial	UNHCR, ACLED, Geospatial
Knowledge	-	SDA, Multiscale simulation approach, File I/O, Acyclic (one-way coupling), hybrid simulation approach	Basic agent parallelisation, agent space parallelisation
Information	-	Secondary	-
Financial	ComPat project	HiDALGO	HiDALGO
Human	Single developer	Team (AJ joined), HiDALGO, Expertise on MUSCLE3 (Lourens Veen from the Dutch e-Science Centre)	Team, HiDALGO
Interpersonal relationships	UCL (ComPat)	Team, HiDALGO	Team, HiDALGO
Structural conditions	Incorporating additional causal conditions through small scale models, enhancing code performance	Effect of weather and route accessibility on distribution of forcibly displaced, Challenges in collection of input and validation data due incorporating additional causal relationships	Evaluate level of detail and visualisation on account of multicausality with implications for computational costs and efficiency
Resource created	Knowledge (multiscale simulation approach), Computational (Parallelised Flee; Coupling file)	-	-

The table above highlights how the resources created during a prototype phase (P5) with limited resources became the basis for multi-specialised functional extensions in the form of multiscale modelling (P10) and scaling and parallelisation (P11). Approaches like coupling were expanded on in P7 and P8 for developing coupling scenarios for food security and conflict evolution. P10 involved collaboration with ECMWF while P11 involved collaboration with the National Technical University of Athens, HLRS, and PSNC within the HiDALGO consortium. Despite resource limitations constraining agency in a prototype phase leading to the development of a simple model, the resources created became the basis for expanding subsequent possibilities, anchored through funded consortia to progressively accommodate the multi-causal nature of forced displacement. These subsequent phases built on the antecedent conditions created in phases like P5 with additional resources made available through funded consortia. This illustrates how combinations of different contextual conditions brought about given outcomes shaping the ongoing innovation process.

Partial resolutions: Partial resolutions, in the form of partial attainment of goals or privileging some goals over others, have been an important aspect of driving the innovation process forward (Sen, 1999; Heracleous & Barrett, 2001; Sabherwal & Newman, 2003). Partial resolutions helps move the process forward when full resolution is not possible. While the ultimate aim of Flee was to help humanitarian organisations in operational planning and preparedness, its successive choices highlighted how it moved towards these phases incrementally through partial resolutions. Table 5 (Resource portfolio) highlighted how agency enabled by the resource portfolio were also subject to resource limitations that constrained them in particular ways. Table 7 (Structural conditions) illustrated the opportunities and constraints presented by multiple environmental drivers that determined the feasibility of particular actions in relation to the hybrid architecture (Table 2) presented by Flee. As a result, as described in Section 5.7, model development proceeds through trade-offs. This included opting for simplicity over scalability in P1 and P5 to arrive at a proof-of-concept in the face of resource limitations. Even when the model development proceeds through widening scope of conditions as illustrated in Table 8 (Scope of contextual conditions) above, it has to contend with resources limitations like data quality (Table 5: Resource portfolio) affected by domain level conditions that lead to revision of numbers and methodological drift in open datasets (Table 7: Structural conditions). These limit the extent to which additional causal conditions can be incorporated within the model and its results appropriately validated. However, incorporating additional parameters and sub-models and evaluating their relationship in terms of sensitivity and performance would require supercomputers making the process computationally expensive and unviable for use within a practical humanitarian context. On the other hand, overly simplified models amenable for use by the humanitarian organisations at the last mile can erode accuracy by making the code more erratic. Moreover, there needs to be a margin of tolerance for error rates to avoid overfitting and ensure wider application. As the model progresses through these trade-offs, it underscores the survival of unmet objectives.

Non-economistic considerations: A contextual and socio-technical approach to theorising admits a diversity of conditions beyond techno-managerial ones. As highlighted in Table 6 (Interpersonal relationships), the ecosystem comprised of a diversity of relationships including extended academic networks, humanitarian organisations, and funded consortia. While the relationship with the funded consortia proceeded through funding and project mandates, significant aspects of Flee's development has proceeded outside of these fixed relationships. The two foundational phases of Flee's development P2 and P3 involved intra-Departmental academic collaborations. Humanitarian organisations as eventual institutional users of the model highlighted modelling requirements and provided critical information resources for rule set updates in P13 and P15 while also shaping an overall understanding of the ground-level dynamics of forced displacement. Similarly, collaborations with the University of Utrecht and Columbia University shaped the model development process by highlighting modelling requirements and granular contextual specificities. These non-financial dynamic relationships existed in tandem with those within funded consortia. In many cases, model development progressed through complementary configurations of these relationships, particularly in phases P7, P8, P14, P13, P15 (Table 3: Actor configurations).

The aim of the research is to understand the interdependent role of the architecture, proximate and distant context within digital product innovation. The two operational research questions disaggregate this relationship to understand the interdependency between the architecture and ecosystem (proximate context) and the nature of the relationship between the architecture, the proximate context, and wider environmental conditions (distant context). The adapted choice framework helps in answering these research questions by providing an analytical approach that helps reconcile the tension between scale and scope of multiple contexts implicated within a given IS phenomenon, in this case digital product innovation process of the Flee ABM model. It helps in explicating design choices for key phases of the process by taking into account reciprocal relationships spanning multiple planes of influence that aggregate into outcomes in each phase. The framework highlights the role of opportunities and constraints that is instrumental in the realisation of a potential choice or outcome that is observed in different phases shaped by the wider set of contextual conditions that act upon it. This helps in understanding how the digital product innovation process evolves through the trilateral interdependency between the architecture and its proximate and distant context i.e. how different conditions of possibility combine and shape collective outcomes not just in different phases but in terms of the overall development of the digital product. It helps negotiate the scope of scale and detail in contextualising the phenomenon of digital product innovation by acknowledging the interrelated role of the architecture, ecosystem, and environment. It further helps go beyond immediate setting of a given outcome to understand the role of operational and enduring choices that structure the digital product innovation process.

6.1.1 Direct and derived conditions of causal trajectory

As highlighted in the theoretical chapter, choice can be attributed to direct and derived reasons. Therefore, analysis must “go beyond the immediate choice of isolated objectives to the emergence and endurance of objectives” (Sen, 1999, p. 272). The preceding section highlighted the scope of contextual conditions directly implicated in the realisation of a given choice and how such choices proceed through partial resolutions as a result of resource limitations, environmental conditions, and nature of relationships. Further, the development process shows the persistence of an enduring choice or a desired outcome and the many operational choices materialised in moving towards such an enduring objective. The enduring choice was for Flee to be operationally relevant to humanitarian organisations, government, and local administrations in helping them pre-position and plan resources and service delivery in conflict driven emergencies that result in forced displacement. Motivated by the 2015-16 refugee crisis, it is tied closely to prevailing policy imperatives around the digital turn and anticipatory action within the humanitarian sector. It is based on forecast driven humanitarian management that aims to mitigate adverse consequences of crises through early action. However, the achievement of this enduring objective requires successive iterative steps to computationally encode and mathematically formulate causal relationships within the phenomenon of forced migration to achieve both the desired level of accuracy as well as economy in terms of computational cost. This highlights the inheritance of this enduring objective in each successive phase and the role of antecedent conditions that help in moving incrementally closer towards achieving it (Sen, 1999; Robey & Newman, 1996). As a result, development proceeds through both direct and derived conditions that helps explain the innovation trajectory. The scope of contextual conditions in each phase are directly implicated in the realisation of its associated outcomes. The scope of contextual conditions is purposively managed through trade-offs or partial resolutions and by leveraging fixed and dynamic interpersonal relationships. Partial resolutions within each phase lead to the persistence of unmet objectives over the course of development. The persistence of such objectives shapes the innovation trajectory in successive phases, for e.g. as illustrated in Table 8 (Scope of contextual conditions), while P5 might have privileged simplicity over scalability or detail, subsequent phases focused on testing scalability through parallelisation to probe the efficiency of the code (P11) or incorporated additional causal factors through multiscale modelling (P10). This further included testing sensitivity of parameters to the output (P13) or increase in the level of detail by incorporating additional parameters through coupling (P7, P8, P10).

Each phase within Flee's development process highlighted the realisation of an operational choice brought about by optimal configurations of resources and relationships. These configurations are drawn from the resource portfolio and relationships that are leveraged from the ecosystem (Appendices 1 and 2). They are in turn implicated in trade-offs or partial resolutions through which the innovation trajectory unfolds (Sen, 1999). This is because given configurations which lead to particular functional extensions involve contending against persistent trade-offs between simplicity and detail, detail and computational cost, flexibility and usability, simplicity and scalability, adaptability and accuracy, and accuracy and flexibility as a result of resource limitations and structural conditions (Table 5: Resource portfolio and Table 7: Structural conditions). This also involved attendant tensions between different resource categories where knowledge resources could be augmented by additional human resource expertise but constrained by computational and data resource limitations in translating social reality into a coded model. Further, information resources which helped highlight additional causal conditions take a long time to be incorporated as rule sets given extant data limitations to model the required relationships. However, successive phases helped take the model a step closer towards realising the enduring choice by either incorporating additional causal details through coupling or enhancing the reliability of the model through sensitivity analysis. These were made possible in successive phases due to the existence of fixed and dynamic relationships within the ecosystem and non-economistic considerations based on reciprocal interest beyond financial incentive. This helped leverage such relationships and even forge complementary alliances (Table 6: Interpersonal relationships) for particular design choices even outside a funding mandate. These included collaborations with humanitarian organisations and academic institutions. Some of the key phases involved development of approaches or creation of resources that widened the ambit for further innovation to take place by resulting in antecedent conditions. For example, the development and incorporation of FabFlee (P3) and SDA (P2) shaped the subsequent innovation process for Flee by enabling automation and replicability through coupling (P7, P8, P10), sensitivity analysis (P13), and MOO (P14).

Understanding and appreciation of the multicausal nature of the phenomenon being modelled along with resource limitations in arriving at its accurate representation meant isolating and modelling causal conditions at each phase by selecting particular configurations of resources and relationships to facilitate particular outcomes. While each operational phase demonstrates a step towards the enduring objective, they are also characterised by trade-offs which are negotiated within the attendant opportunities and limitations imposed by the different resources, relationships, and environmental conditions. Moreover, resource limitations further constrained the extent to which operational choices could move towards the realisation of unmet objectives. These considerations were negotiated by arriving at partial resolutions while working within project mandates and the wider ecosystem where the strength of relationship at a given point in time and dominant environmental drivers shaped their reconciliation towards a given operational choice through the scope of contextual conditions for a given phase.

The foundational phases created the inheritances for the subsequent phases and shaped the nature of the architecture. Hybridity of the architecture was leveraged in multiple ways in the constant move towards realisation of the enduring choice. Its more generic components like FabFlee (P3), created during the development process, became antecedent conditions for subsequent phases enabling further development (Appendix 2). It shaped the way particular relationships within the ecosystem could be harnessed to a given research problem to expand and extend Flee. The research problems emerged through the in-depth investments in understanding each conflict, conversations with humanitarian organisations, and flashpoint events. They were enabled by open source software environments, limited by sectoral data availability, structured by funding inequalities across conflicts, and shaped by extant policy, legal, and regulatory conditions (Table 4: Environmental Drivers and Table 7: Structural conditions). Funded projects contributed to successive availability of resources and relationships by providing financial support for the core team and fostered the freedom to explore research questions aimed at moving closer towards enduring choice, facilitating the availability of pooled expertise, and making scarce computational resources available in the form of supercomputers. This helped stabilise the availability of resources for successive phases that helped in the realisation of different research objectives shaped both by an intuitive understanding of conflict dynamics as well as feedback from the ground. It helped leverage information resources when they became available and translate them to tangible outputs. For example, funded projects also helped contribute to the development of modular toolkits like VECMAtk and SEAVEAtk with multisectoral applications which helped it augment other collaborations by integrating their functionalities, for example, the use of VECMAtk in the development and implementation of the SDSD approach that coincided with the update of the Flee rule set as Flee 2.0 based on feedback received through IOM from frontline NGOs (P13). These built on the SDA (P2) and enhanced Flee's capacity to identify sensitive parameters and improve reliability of the model in future conflict contexts. This illustrates the circular effect of utilisation and creation of resources which eventually became a part of the opportunity structure for successive development. It also highlights the dependency of successive opportunity structures on previous phases and determined how environmental conditions were managed and extant relationships within the ecosystem harnessed through the hybrid architecture which created both opportunities and constraints as a result of its hybrid nature.

Tracing Flee's development highlights a deliberative – evolutionary process of digital product innovation as a result of direct and derived conditions that determine its innovation trajectory. The pathway towards enduring overarching objectives through different operational phases paves the way for evolutionary selection due to the nature of digitality and successive inheritances acquired by the given digital product over the development process. The process is deliberative with the ecosystem mediating purposive selection of complementary resource-relationship configurations within attendant environmental conditions. It is evolutionary because the innovation process itself creates resources while also resulting in unmet objectives that provide the evolutionary conditions for such resources to be taken up in successive phases. The deliberative – evolutionary process highlights how environing conditions are transmuted into the ongoing process of socio-technical digital product innovation whereby the technical and social aspects stand in mutually reinforcing circular relationships that need to constantly negotiate the tensions between technical and social interdependencies. The following sections unpack these relationships in greater detail to understand how such relationships unfold within and subtend the digital product innovation process.

6.2 Coevolution of architecture and ecosystem

The first operative research question tries to understand the coevolution of the architecture and ecosystem in complex products like the Flee ABM. Complex digital products like ABM, unlike modular architectures, exhibit non-standardised design arrangements that organise the mode of its innovation while digitality provides the generative opportunity for collaborative development among a diverse range of stakeholders (Hobday, 1998; Hobday et al., 2000; Yoo et al., 2010; Zittrain, 2006). This question attempts to explore the way complex and non-modular product architectures like ABM structure its development ecosystem and are in turn shaped by the dynamics therein i.e. how the technical architecture determines the way actors and resources converge in particular configurations and how these configurations contain inheritances of the contexts that they are part of which go on to shape the deliberative – evolutionary process of innovation. The Findings showed that an ABM model like Flee displays a hybrid architecture (Table 2: Hybrid architecture) i.e. it maintains its integral character while being supported by generic supportive software and replicable approaches. This hybridity has implications for the way its development ecosystem is structured, organised, and anchored. This is because product architectures provide the scheme by which functionalities of products are assigned to its components (Ulrich, 1995) and thereby provide the method of partition and distribution of design and development tasks within the ecosystem (Sanchez & Mahoney, 1996). As a result, they determine how innovation comes to be organised. Building on multiple inheritances from its previous phases, the generic components like FabFlee allow for automation, inclusion of additional details and modelling additional parameters by coupling them with the original Flee code while its integral character requires deliberative complementary configurations of resources and relationships that shape multilateral interdependencies.

Tightly coupled and integral components are said to stifle innovation by reducing interfaces whereby loosely coupled components can be combined and recombined (Yoo et al., 2010). However, a hybrid architecture exhibits tight coupling with different component characteristics, parts of which have high degrees of reusability and repurposability further augmented by the generative potential of the digital architecture and low marginal cost of recombination and reproduction (Faulkner & Runde, 2009; 2019; Zittrain, 2006). While generic components open up the possibility of innovation, the integral component ensure innovations are closely coupled with given assumptions and parameters about the reality that is being modelled. Changes to the Flee code itself results from how the architecture structures the ecosystem and evolves through complementary resource – relationship configurations. Successive configurations, while resulting in the realisation of a give operational choice, also highlight how the model progresses through partial resolutions in moving incrementally towards its overall objective of being operationally relevant in a humanitarian crisis with high accuracy. The hybrid architecture and its development ecosystem coevolves through complementarity between resource – relationship configurations and operational choices within partial resolutions. The analysis of this relationship proceeds from an understanding of a co-dependent relationship between the architecture and the ecosystem.

6.2.1 Complementary resource – relationship configurations

As a result of the hybrid nature of the architecture, innovation proceeds through successive multi-actor collaborations and combination of resources. Uneven co-occurrence of resources, resources limitations, and existence of both fixed and dynamic relationships (sometimes outside of a funding mandate) necessitate resources and relationships to be aligned in complementary arrangements that can result in the materialisation of given outcomes. Consider the case of automated location graph construction (P12). P12 brought together the team and HiDALGO consortium partner KNOW Centre to not only pool shared resources within the HiDALGO project but also leverage access to supercomputing resources of the Austrian COMET programme of which the Centre was a part. Similarly, while the facility location problem (FLP) did not find subsequent purchase after P4 due to lack of reciprocal interest, it was included as a part of P14 for the simulation – optimisation approach to identify optimal camp location. This came to be combined with computational resources created in P12 in the form of the route pruning algorithm along with the project mandates for HiDALGO and ITFLOWS which provided both the funding mandate for the functional extension as well as the terminology for legal harmonisation.

The hybrid architecture defines rather than divides tasks in a way that maximises the overall functionality, accuracy, and relevance of a product. Its task definition derives from the deliberative – evolutionary process of innovation whereby its enduring choice of relevance in humanitarian action can be broken down to the need for accurate prediction, determination and modelling of additional causal parameters, and improving the usability of the model. The development of the Flee code requires specialised knowledge in terms of its overall design. Given the high level of system awareness, the team as system integrator is able to define initial prototypes, the design problem (initially intuitively, then through combination of secondary sources and contribution of ecosystem participants like humanitarian organisations). Hybridity of the architecture helped create conditions of possibility for successive digital innovation initiatives where the main code could be coupled with additional details to expand the scope of causal conditions being modelled like in P7, P8, P10.

Ecosystems around complex digital products with hybrid architectures are not neatly partitioned into roles and tasks. Rather than segmenting the ecosystem into stratified roles, the hybrid architecture structures the ecosystem on the basis of specialisations which are then brought together in configurations depending upon the design problem to be resolved. As described in section 5.3, the team itself contained members with diverse specialisations. The HiDALGO consortium brought additional specialisation on high performance computing and data analytics while VECMA brought in specialists in computational mathematics to develop VVUQ approaches to evaluate sensitivity of the output to given parameters. Similarly, the ITFLOWS consortium brought together a multidisciplinary team including lawyers who advised on appropriate terminologies to ensure legal harmonisation and validity of the outputs. These specialised expertise is enabled by the agency conferred by the resource portfolio and constrained by resource limitations and their unequal co-occurrence (Table 5: Resource Portfolio).

The successive multi-actor alliances with diverse specialisations are in turn shaped by fixed and dynamic relationships which help bring together this required expertise and distributed resources for development of a particular aspect of the architecture. Fixed relationships involved those defined by project mandates while the dynamic ones involved those with non-economistic considerations in the form of reciprocal utility like humanitarian organisations and academic institutions. The combination of the nature of fixed and dynamic relationships brings about the complementary convergence of distributed resources beyond narrow self-interest through interpersonal comparisons of reciprocity and collaboration (Sen, 1999). Participants collaborate not just within the funding mandates but also across project boundaries where scientific and operational interests coincide. P13 illustratively combines expertise from VECMA, HiDALGO and a humanitarian organisation (IOM) in developing the SDSD approach which also marked the release of Flee 2.0. Long term relationships with humanitarian organisations, albeit ad hoc, proceed through a mutual understanding of reciprocity wherein multiple inheritance from diverse projects are aligned with modelling requirements provided by them. Stakeholders and resources come to converge in complementary configurations which further subdivides a given task according to specialisations among participants in a given phase.

These complementary resource – relationship configurations both shape and are shaped by how hybrid architectures structure its development ecosystem and is in turn shaped by its multilateral inheritances derived from multiple planes of influence. The nature of the architecture helps organise the configurations of resources and actors through a focused design problem that can help realise the expected outcome of a given phase (Hobday, 1998; Hobday et al., 2000; Adner, 2017). The hybrid architecture does not contain a neat partition between its constituent parts but involves a tightly coupled technical dependency that both requires and facilitates the participation of a diverse range of stakeholders while also structuring the nature of their participation in terms of what this technical interdependency will admit. The hybrid nature of the architecture creates both opportunities and constraints for a multilateral set of partners to converge in the realisation of operational and enduring choices. While each phase saw the coming together of a given set of partners (e.g. Table 3: Actor configurations and Table 6: Interpersonal relationships) for its operational realisation, these relationships could not be reduced to bilateral ties alone (Jacobides et al., 2018) due to diversity of resources required in each phase which were brought together by the range of stakeholders that comprised the development ecosystem of the product. Participation in each design phase is not unlimited and unfettered but curated, structured and selective that aligns complementary set of resources that generate a particular outcome within a given phase. While the team as the system integrator is able to perform the search and selection to bring about the necessary resource – relationship configuration, its ability to do so is structured by the constraints and limitations among a set of given resources necessary to bring about a given innovation effort.

These complementary resource – relationship configurations that bring about an outcome within a given operational phase is representative of how they helped in achievement of given operational choice. For example, the integrated IDP forecasting phase (P15) saw the team leveraging its relationship with Save the Children and information provided by them with the qualitative reports from secondary sources and open data (UNHCR, ACLED, and OpenStreetMaps) under the global challenges aegis of HiDALGO. Yet it had to contend with resource limitation in the form of patchy data availability compounding the process of incorporating the richness of information resources within existing computational and mathematical knowledge resources (Table 5: Resource portfolio). This is due to the time intensive process of harmonising and integrating the complex causal pathways for both refugees and IDPs into machine interpretable computational and mathematical causal links some of which the team is attempting to capture in Flee 3.0. Similarly for multiscale modelling and weather coupling (P10) and SDSD approach (P13) where the need to extend relevance and reliability of the model saw leveraging of fixed and dynamic relationships from funded consortia and humanitarian organisations.

However, the evolution of these functional advancements often does not highlight the background conditions where digital innovation processes were hindered or constrained due to bottlenecks. For example, Iranian postdoctoral researchers within the team were unable to use supercomputers at HLRS within the HiDALGO project because they were made in the United States and therefore came under legal sanctions imposed by the US on certain nationalities (second review feedback). Moreover, the persistence of deliberative resource – relationship configurations into evolutionary conditions would only unravel over time and was dependent on successive complementary alignments which was an ongoing process.

As discussed above, when certain resource categories like data structure the limitations for how the model operates, these limitations are managed through partial resolutions and complementary combinations with other resources and relationships within a given phase. For example, sensitivity analysis helps guide simulation development and refinement avoiding direct calibration to validation data. These limitations and complementary configurations help in defining the trade-offs that shape the model development process and its trajectory towards realisation of an operational choice. For example, using a detailed SDSD model helped evaluate sensitivity of parameters but was only possible using available computational resources like supercomputers to refine and streamline the model going forward or performance benchmarking on supercomputer to test efficiency of the code. These examples highlight how despite detailed models raising computational cost, they were deployed when computational resources were available to identify conditions for their eventual simplification and efficiency. Partial resolutions in a given phase lead to the survivability of unmet objectives which are taken in up subsequent phases, e.g. the first prototype versions such as P1 and P5 prioritised simplicity over detail and scalability. Subsequent phases like P7, P8, and P10 worked towards iteratively including additional levels of details in the form of food security, conflict evolution, and multiscale modelling.

Since adding additional levels of detail can increase computational costs, it leveraged supercomputing capabilities from extant funded consortia in P8, P10, P11, and P13. Similarly, as discussed earlier, P11 extended the output of P5 and went on to prioritise scalability of the algorithm to assess efficiency.

The hybrid architecture was composed of resources created during the innovation process. It involved high interdependency among technical components that structured its evolutionary trajectory. While components are easily identifiable, their decomposed units cannot deliver the overall functionality of the code. The overall behaviour of the model remains sensitive to its input parameters and assumptions while the more generic components exist in supportive functions to the integral component i.e. the main Flee code (Table 2: Hybrid architecture). The foundational phases of the initial model (P1), generalisation (P2), and automation (P3) that resulted in the creation of computational and knowledge resources like FabFlee and SDA helped facilitate ecosystem formation and emergence. The foundational phases established the prototype, foundational tools, and approaches that helped in grant applications for funded projects which established the basis for fixed relationships within the ecosystem shaped around project objectives like HPC, HPDA (HiDALGO), VVUQ (VECMA), and large scale migration prediction for the EU (ITFLOWS). However, the formation of ecosystem around product development also proceeded from the larger professional networks which provided the linkages and introductions to formalise relationship through collaboration and concretising existing looser academic ties into stronger relationships under funded projects (Table 6: Interpersonal relationships). The foundational phases that resulted in the development of prototype also helped in forging dynamic relationships with humanitarian organisations and resulted in modes of collaboration contributing to the overall behaviour of the model through informational resources that resulted in rule set updates.

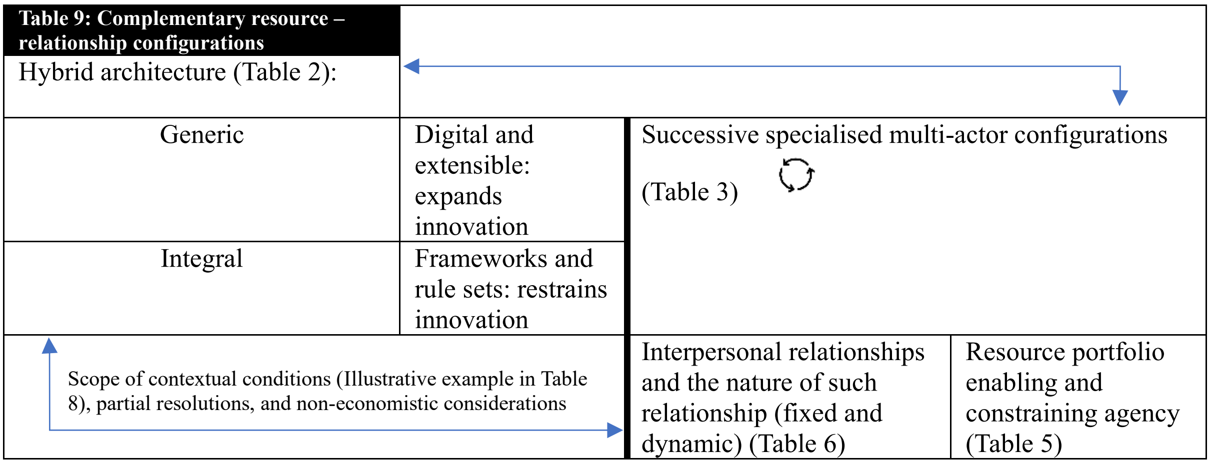
Generative potential of existing computational resources was harnessed through knowledge and human resources by developing and integrating FabFlee (P3) with the original Flee code while the process was made more replicable through the SDA (P2), moving beyond the original hard coded 2016 model (P1). FabFlee then became an important resource component in successive opportunity structures anchoring the development of solutions to different modelling considerations. Thereby, becoming the space not only for revisable configurations (Alaimo et al., 2020) but also enabling different resources and relationships to coalesce within the ecosystem to shape the outcome within a given phase of innovation. While generic components yielded opportunities to anchor development within ecosystem objectives, the integral nature of the core Flee code itself necessitated the formation of dynamic linkages with academic and humanitarian stakeholders that could provide the relevant information for rule set updates. Rule set updates through information resources were aimed at bringing the behaviour of the model closer to reality, by incorporating rest stops, updating movement speeds, and / or explicating travel conditions, enhanced through integrated sensitivity analysis (P13). Incorporation of information resources through rule set updates helped bridge the limitation of unreliable data sources and the inability of computational and mathematical knowledge resources in the capturing the complex multicausal phenomenon that is forced displacement. While funded projects provided the sustainability for experimentations to continue, collaboration beyond funded initiatives involved moving towards enduring choice by approximating reality more closely. The combination of complementary objectives from funded projects and unfunded collaborations also helped in the incorporation and computational rendition of evolving requirement specifications as transpired during rule set updates. Thus, while generic modular components provided the space for multilateral collaboration, funded projects provided the basis for stable evolvability.

The stability – evolvability dichotomy is an enduring tension within ecosystems by affiliation that needs to be actively managed for coevolution of architecture and ecosystem (Wareham et al., 2014). However, ecosystems around hybrid architectures require simultaneous conditions of stability and evolvability to be present in order to coevolve. Stable evolvability can be described a condition whereby the financial resources provided by funded projects offered sustainability for the team to maintain the product development process. Evolution within the process was ensured by enabling successive opportunity structures of resource – relationship configurations through alignment of complementary resources, project objectives, and requirements and expertise from fixed and dynamic relationships as was observed from P7 onwards. For example, the VECMA and HiDALGO projects provided the financial resources to leverage dynamic relationships with IOM and Save the Children to incorporate rule set changes in the form SDSD approach (P13) and exploring the IDP modelling and forecasting (P15). When information resources and requirements from humanitarian organisations were aligned with project objectives, it helped leverage shared project resources in the form of human resource expertise and scare computational resources like supercomputers through HPC and sensitivity analysis thereby helping to resolve some of the extant unmet objectives. Stable evolvability was also ensured by project outputs in the form of computational resources like generic software toolkits like VECMAtk and SEAVEAtk which have general purpose application. They represent the modular components to the extent that they can be used and reused in different application scenarios by providing a replicable computational environment.

Stable evolvability enabled the team to continue their search and learning by collaborating with user stakeholders like humanitarian organisations which yielded modelling requirements in use. It helped in incorporating both requirements specifications and information resources that were instrumental in moving the model a step closer towards the conception of the reality it tries to represent. The combination of financial resources from different projects also necessitated the alignment of different aims and objectives of multiple research projects to develop diverse functionalities within Flee in a homogeneous composition. Each actor within the development ecosystem operated with different objectives and their collaborative involvement within the development process involved successive stages of development with multilateral inheritances from each actor group. The collaborations were enabled not just through narrow self-interest but through shared objectives of the scientific consortia, the complementary configuration of specific expertise, and dynamic relationships of reciprocal utility like in the case of humanitarian organisations, University of Utrecht and Columbia University. Therefore, the Flee code was able to progress through various operational stages towards its ultimate objective by incorporating requirement specifications from its user groups by integrating it with project mandates. These multilateral inheritances meant that functional extension under one project could be augmented under another, for example, sensitivity analysis was augmented through HPC computing. HPC environments helped test the performance and scalability of detailed simulations as well as perform sensitivity analysis which could help in identifying the most relevant parameters for a more frugal version of the model to optimise computational cost and ensure relevance in under-resourced settings.

As highlighted above, digital product innovation through coevolution of architecture and ecosystem occurs through complementary resource – relationship configurations which requires holistic overview of the innovation process with regard to a given task, alignment across different relationships, and overall development and evolution of the product. Ecosystem around hybrid architectures like ABM does not involve a stable core and heterogeneous complementors working towards maximising variety and variability of products (Baldwin & Clark, 2000; Yoo et al., 2010). The architecture and ecosystem coevolves through selection and convergence of resources with regard to a particular problem space, multilateral inheritances, and resource limitations. The ecosystem is characterised by a diverse set of stakeholders with high user involvement (Hobday, 1998; Hobday et al., 2000) where such users like humanitarian organisations both provide information resources and specify modelling requirements due to non-financial reciprocal relationships as highlighted in section 6.1. As much as coevolution occurs through the way the hybrid architecture facilitates participation in a particular way, it also occurs through the configuration of resources within opportunity structures and multilateral relationships within the ecosystem. Sometimes when a given configuration has brought about a particular development, background conditions at play (such as in the form of US sanctions against Iranian nationals) limit the scope of contextual conditions even though the overall the model appears to be on its evolutionary trajectory. Therefore, as summarised in Table 9 (Complementary resource – relationship configurations) the architecture and ecosystem co-evolves as the hybrid architecture (Table 2) organises the ecosystem through successive specialised configurations of ecosystem participants (Table 3) to solve particular design problems through interpersonal relationships and attendant resource portfolio and associated limitations. These needs to be aligned in complementary configurations to create antecedent conditions to drive the innovation process. This requires the purposive configuration of the scope of contextual conditions, management of trade-offs or partial resolutions given the hybrid nature of the architecture and resource limitations, and nature of relationships to be leveraged and aligned.

Table 9: Complementary resource – relationship configurations



Hybrid architecture (Table 2):			
Generic	Digital and extensible: expands innovation	Successive specialised multi-actor configurations (Table 3)	
Integral	Frameworks and rule sets: restrains innovation		
Scope of contextual conditions (Illustrative example in Table 8), partial resolutions, and non-economistic considerations		Interpersonal relationships and the nature of such relationship (fixed and dynamic) (Table 6)	Resource portfolio enabling and constraining agency (Table 5)

6.3 Interrelationships between environmental conditions, architecture, and ecosystem

Wider environmental conditions are also implicated within complementary resource – relationship configurations driving the innovation process. The Findings highlighted role of four categories of environmental drivers: computational, digital, and data; domain; socio-political conditions; and policy, legal, and ethical. These environmental drivers shape the problem space, opportunities and limitations of existing resources, demand conditions, and legitimacy. Many of the resource limitations highlighted in Table 5 (Resource portfolio) like data quality are linked to wider domain level environmental drivers like funding inequalities and lack of granular data as highlighted in Table 7 (Structural conditions). Similarly dynamic ad hoc relationships like the ones with humanitarian organisations tend to reignite as a result of socio-political conditions like flashpoint events.

Unlike traditional product development, the optimisation of ABM does not aim for a full contextual fit (Clark, 1985) because overfitting to the particular context will make them unsuitable for others. This highlights the role of the distant context of the wider environment both in terms of provision of material resources as well as setting the conditions which structure the development and innovation process (Constaninides & Barrett, 2006; Zucker, 1983; Jones, 1999). While these conditions originate beyond the immediate setting of Flee's development, they come to determine the development of the product and its proximate context.

The role of the wider set of environmental conditions helps situate the digital product innovation process within the larger context within which it operates. While the co-evolutionary dynamic of the architecture and ecosystem helps explicate the role of conditions within its proximate context, situating them within wider environmental conditions help analyse the interrelationships and enable multi-context analysis. As highlighted through Table 4 (Environmental drivers), different environmental drivers work in complex circular relationships creating demand conditions through flashpoint events, placing resource limitations like availability of reliable data, or enabling opportunities through open source computational resources. Policy and legal drivers can both provide impetus for and drive the model towards greater legitimacy through legal harmonisation.

The wider environmental context and the range of drivers therein helps to understand the multiple interconnected levels of contexts and reciprocal relationships therein (Avgerou, 2019; Pettigrew, 1997). Understanding the interdependent role of the three constituent elements simultaneously helps acknowledge the interrelationship between the social and technical components that frame the digital product innovation phenomenon. This also helps take into account the layered relationship between the distant and proximate context by throwing into sharp relief how conditions of possibility crystallise into direct conditions leading to outcomes for given phases. It highlights the pathways through which the team as the system integrator with overall specialised knowledge of the products works to navigate the opportunities and constraints placed by the ecosystem and environment in which it is embedded. It highlights the trade-offs that characterise how operational choices are realised in moving towards the overall achievement of the enduring choice. This helps explain how relationships develop, converge, and sustain in the context of these intersecting and multidirectional influences with the need to highlight the nature of this trilateral interdependency that shapes these concerns.

6.3.1 Ecosystem as a mediating construct

Multiple environmental drivers stand in inter-locking relationship with the architecture and its proximate context due to opportunities and constraints derived from multiple contexts that need to be managed through resource – relationship configurations. The previous section highlighted how such configurations come to be structured by resource limitations and the multilateral inheritances that different stakeholders bring to the ecosystem. It highlighted how such configurations moved towards stable evolvability by leveraging the open source environment where the prototype helped attract additional resources for further development thereby shaping the conditions for their evolution. Within the deliberative – evolutionary innovation process, the ecosystem mediated and managed the multidirectional influence presented by multiple environmental drivers. As illustrated in Table 8 (Scope of contextual conditions), in the initial stages of development, resources beyond open source computational and data resources and modelling knowledge, access to other resources like information, finance, and human resources were limited. As the prototype advanced through the foundational phases, it was able to attract additional resources required during each phase. However, given the nature of dynamic and fixed relationships their configurations across given operational phases differed. This is because for example, while the relevant human resource expertise could be managed through fixed relationships, information resources that determined the behaviour of the model through rules sets updates depended on dynamic ones. While expertise for coupling additional parameters (P7, P8, P10), automating processes (P3, P9, P12), developing approaches for sensitivity analysis (P9, P13) or incorporating policy and legal advice (P14) could be leveraged through the specialists within the team and funded consortia, major rule set updates required multi-actor coalitions to be complemented by humanitarian organisations (P13, P15). As highlighted in Table 5 (Resource portfolio), mathematical and computational knowledge resources were always constrained in limiting the extent of the complex social reality that could be computationally or mathematically rendered. Domain level specificities like conflict dynamics and terrain and weather conditions limit the extent to which a completely replicable model can be developed without running the risk of overfitting. These conditions also highlighted the need to drive the innovation process forward through partial resolutions in each phase in the light of attendant opportunities, limitations, and constraints.

While the team purposively curated complementary resource – relationship configurations across different operational phases, it did not possess most of the necessary resources required during the innovation process beyond open source software, open data, and some diversity in modelling knowledge. The deliberative search and combination of these resource – relationship configurations help in driving evolutionary selection through the creation of antecedent conditions in the form of computational and knowledge resources and maintaining dynamic relationships even in the absence of financial incentives as highlighted in Appendices 1 and 2 and Section 6.1. The lack of control over necessary resources that facilitate stable evolvability of the ecosystem means the team exercises its system integration role in way that enables the ecosystem to manage and mediate both the limitations of its role, available resources, and the environmental pressures in continuous effort to find relevance and legitimacy. By maintaining such a role it is able to leverage resources and relationship from the extant ecosystem in response to sudden onset environmental conditions in the form of flashpoint events like the outbreak of war or civil conflict (a socio-political driver). This also helps harness opportunities in the environment in the form of open data and open source software and select the appropriate parameter to model within a particular conflict context since as a result of the multicausal nature of forced displacement all causal relationships cannot be computationally or mathematically rendered. It also helps leverage and navigate the extant policy and legal environment through diverse expertise within its ecosystem.

Therefore, the ecosystem acts as a mediating construct in both leveraging opportunities and managing constraints within the environmental context in which it is situated while working within the mandates that govern its fixed relationships and the conditions that shape its dynamic ones.

This highlights how the conditions created by these multiple environment drivers are transmuted, translated, or mediated into outcomes and action through the ecosystem. For example, ethical and legal concerns could be addressed through fixed relationship within ITFLOWS (P14). While socio-political drivers like flashpoint events (Table 4: Environmental drivers) create demand conditions to model new conflict, domain level specificities within new conflict conditions as in the case of Tigray, elicit fundamental rethinking around rule sets to include IDPs (P15). This circular, reciprocal, interdependent relationship highlights the mutual shaping of the social and technical components. Environmental conditions provide the initial conditions for the foundational phase through availability of open sources software that are combined by the team into the Flee code and updated through generalisation and automation. These were particularly helpful for developing prototypes as in P1 and P5 while open data like UNHCR and ACLED continued to be the fundamental resource inputs throughout the life of the development process in the absence of alternatives.

However, in order to continue the development process, the stable evolvability of the proximate context of the ecosystem was required to both respond to and negotiate the combined impact of multiple environmental drivers and attendant considerations of the hybrid architecture. Therefore, while the environment provides both the opportunities and constraints for the development of the prototype through multiple environmental drivers, their materialisation through the innovation process depends upon the extent to which generativity is able to circumscribe or approximate the given range of reality given resources limitation posed by the very same drivers.

This requires an entity that mediates and negotiates these conditions. As a result, the team coordinates extant relationships within the ecosystem in different configurations over different phases to drive the innovation process forward. Given multilateral relationships and resource limitations, the innovation process proceeds through partial resolutions where the operational choice at each phase underscores the survival of unmet objectives necessitating the ongoing negotiation of attendant limitations, constraints, and opportunities. The team at the first instance mediates the environment into the first prototype and foundational phases which results in a hybrid architecture which enables it to create the foundations for an ecosystem to emerge by leveraging different relationships. The ecosystem then becomes the mediating entity that is continually engaged in the translation from architectural, resource, and environmental conditions to outcomes. While the resources themselves shape the ability for the team and its network of stakeholders to realise a given choice, they do so within the bounds set by the environmental conditions and the hybrid architecture. Thus, the ecosystem becomes instrumental in straddling this negotiated innovation process. As highlighted in Table 3, the ecosystem is characterised by successive multiple actor configurations. The diversity and nature of interpersonal relationships therein determine the collective agency that can be exercised at a point in time based on resource availability (as illustrated in Table 6: Interpersonal relationships). Table 9 (Complementary resource – relationships configurations) highlights how the architecture and the ecosystem co-evolves with the hybrid architecture organising the ecosystem into multi-actor specialisations whose successive configurations help in solving focused design problems in relation to the architecture. However, the ability of these configurations to realise given outcomes are determined through the purposive management of the scope of contextual conditions, partial resolutions, and nature of relationships in the light of resource limitations and hybridity of the architecture which in conjunction with structural conditions shape the feasibility of particular actions which come to be negotiated through the ecosystem within and across phases. Therefore, as highlighted in Table 10 (Ecosystem as a mediating construct) below, the ecosystem negotiates the potential possibilities and limitations and opportunities and constraints arising out the nature of architecture, different environmental drivers and resources through management of successive complementary resource – relationships configurations driven by direct (implicated within a given phase) and derived conditions (resource creation with subsequent uptake and survivability of unmet objectives due to partial resolutions).

Table 10: Ecosystem as a mediating construct

Hybrid architecture (Table 2):	Ecosystem mediates through management of complementary resource – relationship configurations	Innovation proceeding through direct and derived conditions
Restrained expansion of innovation		
Environmental conditions (Table 4):		
Opportunities and constraints (Table 7)		
Resource portfolio (Table 5):		
Agency and limitations		

6.4 Role of technical and contextual interdependencies in digital product innovation

The thesis aimed to look at socio-technical interdependencies between the architecture and its proximate and distant context within digital product innovation for complex products like Flee ABM. The above two operational research questions explored the role of the overall context by disaggregating it into proximate and distant context of the ecosystem and environment respectively. While it acknowledges the role of multiple contexts that lead to multilateral inheritances in each of these contextual spheres, the demarcation helps understand combination of complementary forces in order to develop holistic explanations. The above sections highlighted the circular mutually reinforcing relationships between architecture, ecosystem, and environment in shaping the process of innovation. It demonstrated the role of proximate environment in providing the immediate conditions for deliberative aspects of the innovation process. However, these deliberative actions also lead to conditions for evolutionary selection through the creation of computational and knowledge resources that are leveraged in resource – relationship configurations in successive phases and the partial attainment of objectives leading to the survivability of unmet objectives over time.

The generative system that is built up through combination of technological arrangement of digital tools among groups of actors (Zittrain, 2008) is a product of negotiations between social and technical conditions between the architecture, the proximate, and distant context that make up generative systems and its evolution. The constant negotiation of opportunities and constraints conditions the generative potential of these systems as such potential comes to be bounded by interlocking multi-contextual constraints. This comes to be shaped by the layered-relational nature of this trilateral interdependency where proximate context mediates the layered structural conditions of the wider environment through relational dynamics within it.

6.4.1 Bounded generativity – A contextualised approach

Inherent generative potential of digital products is a key driver for innovation as it enables endless recombination (Yoo et al., 2010; Zittrain, 2006; 2008; Benkler, 2006). However, interdependencies between a digital architecture and its proximate and distant context bounds such generative potential through direct and derived conditions of the deliberative –

evolutionary process of complex digital product innovation. Table 9 (Complementary resource – relationship configurations) showed how resources and relationships need to be aligned in complementary combinations within given phases wherein an outcome is brought about through the scope of contextual conditions the team as a system integrator purposively brings together by leveraging fixed and dynamic relationships, entailing trade-offs and partial resolutions. Further, as demonstrated in Table 10 (Ecosystem as a mediating construct) these complementary configurations within the ecosystem negotiate attendant opportunities, possibilities, constraints, and limitation of the architecture, resource portfolio, and the environment that shapes the feasibility of action wherein innovation comes to proceed through the direct and derived conditions. Thus, bounded generativity refers to the way ecosystem mediates environmental conditions within hybrid architectures that places boundaries upon the extant generative potential of digital products and how it is transmuted into operational outcomes while moving towards enduring overarching objectives of the digital product innovation phenomenon. This notion of bounded generativity as a contextualised approach and a deliberative – evolutionary mode of digital product innovation stands in contrast to the notion of bounded generativity proposed by Fürstenau et al. (2023) as the stabilisation of ecosystem boundaries. Privileging the ‘product view’ of generativity, it highlights how the extant opportunities and constraints in the architecture and environment are negotiated through the ecosystem which places limits on product boundaries by guiding the direction of product development through purposive *ex ante* negotiations based on successive complementary resource – relationship configurations. This mutually shaping trilateral interdependency highlights how the extant ecosystem that comes to be organised around the hybrid architecture relates to guided direction of product development rather than stabilisation of third-party affiliation within a platform ecosystem.

The ABM digital product innovation process moved through successive complementary resource – relationship configurations within the proximate context. The distant context of environmental drivers present interrelated conditions that stand in inter-locking relationships with the architectures and the resources that shaped such resource – relationship configurations. As highlighted in Table 7 (Structural conditions), environmental drivers presented opportunities and constraints which often worked in tandem with one another. The existence of the open source environment helped bring about the necessary resource configurations to produce a prototype as a single developer. However, as successive development phases showed, moving beyond the tightly hard coded model towards a hybrid architecture required additional resource – relationship configurations focusing on different aspects of the modelling problem (Illustrative example in Table 8: Scope of contextual conditions). Resource – relationship configurations inherited the layered – relational dynamic between the proximate and the distant context where project mandates, requirements specifications, and resource limitations had to be managed to drive the innovation process forward. Within the ecosystem the complementary resource – relationship configurations shape the proximate context of the innovation process. However, these configurations inherit limitations and opportunities from the distant contexts from which they are derived. These inheritances necessitate trade-offs and partial resolutions which shaped the evolutionary conditions within the development trajectory (Table 10: Ecosystem as a mediating construct; Sections 5.7; 6.1) for example, through the development of computational and knowledge resources which create antecedent conditions, choosing simplicity over scalability, not optimising for error rates to minimise over-fitting, limiting the range of detail to reduce computational cost among others.

While stable evolvability provided by funded projects in the proximate context allowed the necessary stability to explore, expand, experiment and negotiate attendant tensions it also placed boundaries through research objectives to be met as a part of funded project mandates (Table 5: Resource portfolio, financial resources). As a result, negotiating collaborations through dynamic relationships necessitated and required alignment between project aims and speculative explorations based on requirements specifications from user-stakeholders like humanitarian organisations. This also shaped the direction of the Flee is particular ways, for example, the focus on HPC within HiDALGO moved Flee in the direction of parallelisation of the code to enable parallel computing (P11) while sensitivity analysis and VVUQ testing within VECMA led to integration of the same within SDA in the SDSD iteration (P13). The focus on legal harmonisation within ITFLOWS ensured the use of appropriate terminology thereby extending its legitimacy as well integration with a larger global model like EMT (P14; Field notes). However, these projects also provided the conditions to computationally encode and model incoming informational resources in a way such that the behaviour of the model moved to closer to real world dynamics. Despite the proximate context including mandates like HPC testing which are operationally untenable in practice outside the project scope, it helped in evaluating and streamlining the performance of the code for future use.

Specialised complex products like ABM composed of generic and integral components contain layers of technical interdependency that subtend the social ones. Within the technical architecture the more generic supportive software opened up spaces for consequent development while the integral nature of the Flee requires more substantive rule sets updates and calibrations through sensitivity analysis and manual incorporation of information resources and iterative refinement and testing (Table 2: Hybrid architecture). Moreover, overfitting within a particular context would render the model unfit for cross-contextual application at the domain level (P2). Such conditions circumscribe the generative potential of the digital product under development. The proximate context conditions the generative potential through mediations of the distant environmental drivers and architectural conditions. For example, domain level informational inputs transitioned from intuition (P2) to insights from humanitarian organisations (P13, P15). However, the limitations of existing data resources placed restrictions on the extent to which Flee could approximate the full scope of such information provided. Some of these limitations were shaped by environment conditions which determined the extent to which the ecosystem could mediate in responding to the enduring demands placed by it (Section 5.4, 5.6.3, Table 7: Structural conditions; Table 10: Ecosystem as a mediating construct).

One of the enduring trade-offs was the continuously unfolding reconciliation between knowledge, information, and data resources in abstracting the complex multicausality of forced displacement (Section 5.6.1). The systemic issues around data availability and methodological drift underlying open data within the sector undermine data quality. This is compounded by domain level issues like funding inequalities among conflict locations and the lack of appropriate substitutes for data sources like UNHCR (Table 7: Structural conditions; Sections 5.4; 5.6.3). Both data limitation and the extent to which mathematical knowledge resources can model the complex nature of forced displacement limit the extent to which information resources can be fully incorporated within the modelling process. This is because the diversity of causal factors that exist in social reality cannot potentially be mathematically formulated and computationally rendered without destabilising the model and pushing computational costs that would undermine the enduring objective of it being ultimately relevant within a humanitarian management context. While partial resolutions of incorporating them in assumptions and rule sets are undertaken, potential inclusion of the full scope of information reality remained limited. Alongside the deliberative aspect of the process, evolutionary selection is only possible due to the existence of antecedent conditions created in previous phases.

Both proximate and distant contextual conditions place interlocking constraints on the direction of generative potential. The deliberative – evolutionary process of moving towards enduring objectives through partial resolutions, driven by a diversity of proximate and distant conditions, and the hybrid architecture determine the trajectories of the innovation process. Despite the wide range of actors and potential for collaboration from the wider academic and humanitarian community, relationship configurations are shaped on the basis of their complementarity with available resources and modelling requirements (Table 9: Complementary resource – relationship configurations). The environment contained the limitations and conditions of the resources used within the development of the model and also created the demand for the product to remain in existence (Table 7: Structural conditions). The purposive curation of complementary resource – relationship configurations to mediate tensions arising from multiple environmental drivers, resource limitations, and technical considerations of the hybrid architecture circumscribes the indeterminacy of unbounded generative potential and inhibits boundless recombinant development (Table 10: Ecosystem as a mediating construct).

While the generative nature of digitality within the hybrid architecture of the ABM model helped provide locations for stakeholders and expertise to converge, this has to contend with its overall integral composition. This required leveraging fixed and dynamic relationships within the ecosystem and alignment of projects mandates and requirements specifications. Moreover, complementary configurations which enabled the realisation of a given choices were conditioned by the limitation and opportunities posed by different resources and their arrangement in generative combinations which helped in advancing the innovation process. As a result of these different constraints placed by the technical and environmental conditions, the ecosystem served as a mediating construct that led to the materialisation of given operational choices. It highlighted how despite the generative potential which drove the innovation process, different conditions in the proximate and distant environment and the nature of the hybrid architecture itself bounded the generative trajectory in successive operational choices which underpinned the digital product innovation process. Revisiting Fig. 1, Fig. 13 below highlights bounded generativity as an explanatory construct for the deliberative – evolutionary mode of complex digital product innovation in theorising the interdependencies between the architecture, ecosystem, and environment.

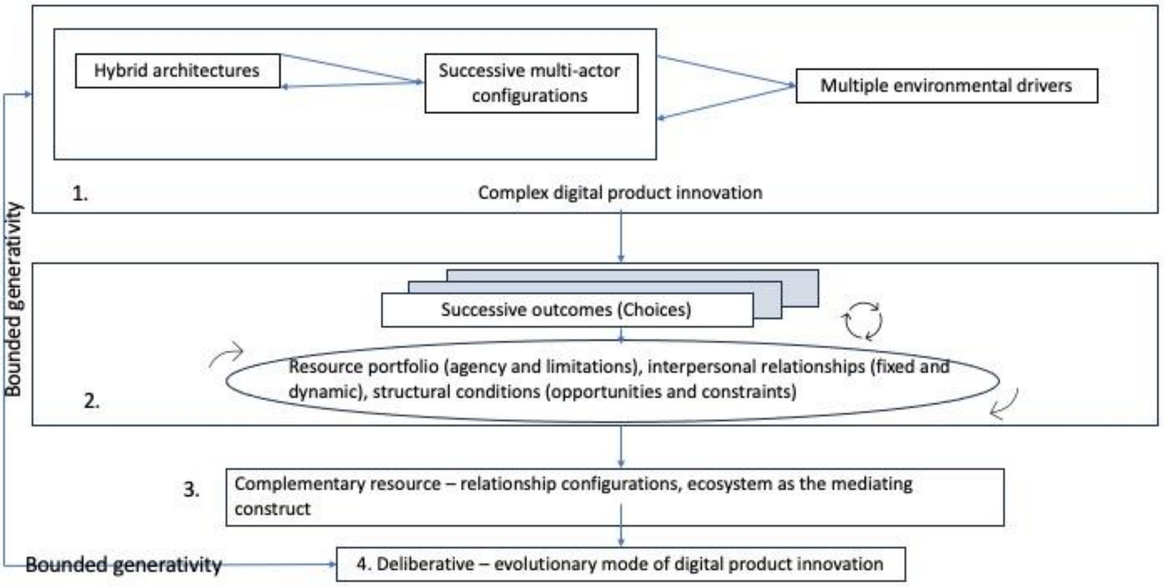


Fig. 13 Bounded generativity – a contextualised approach for digital product innovation

Developing a theory of the complex digital product innovation proceeded through descriptive accounts of the architecture, ecosystem, and environment as identified in Layer 1 of Fig. 1. The hybrid architecture organises the ecosystem in successive specialised multi-actor configurations within the presence of multiple environmental drivers. Innovation progresses through the presence of resource portfolio, interpersonal relationships, and structural conditions which present diverse opportunities and limitations that stand in inter-locking relationships with each other (Layer 2). As a result, the scope of contextual conditions for each phase involves complementary resource – relationship configurations to navigate these attendant conditions. The ecosystem acts as a mediating construct driving the innovation process forward on the basis of direct and derived conditions within and across successive resource – relationship configurations (Layer 3). This becomes necessary due to the survival of unmet objectives due to the nature of the architecture, resource limitations and opportunities and constraints presented by different environmental drivers which shape the feasibility of action within and across phases. As a result, each phase involves the purposive curation of complementary resource – relationship configurations that is directly implicated within a given outcome. Concurrently, resource creation in certain phases with subsequent uptake and partial resolutions in each phase leading to the survival of unmet objectives serve as derived conditions of successive phases. Jointly they shape the innovation trajectory of complex digital products like Flee. As a result, the innovation process comes to be both deliberative as a result of direct conditions in given phases and evolutionary as a result of derived conditions that drive evolutionary selection (Layer 4). This deliberative – evolutionary mode of complex digital product innovation bounds the generative potential of digital products through their constant purposive negotiation of opportunities and constraints which come to characterise the innovation trajectory of complex digital products.

6.5 Summary and synthesis

This chapter foregrounded the nature of the digital innovation process and how such a process is determined by the nature of interdependency between the architecture, ecosystem, and environment. Particularly how such relationships unfold around architectures that do not conform to modular design principles that have underpinned IS theorisation on digital product innovation where aspects of scale and mass market conditions determine the ongoing development of the digital product. While the study is premised on the existence of relationships between the architecture, ecosystem, and environment, this chapter demonstrates how the nature of such relationships shapes the digital product innovation process. With the aim of constructing causal explanations behind the phenomenon, this chapter submits that digital product innovation process for complex digital products like ABM have to be attentive to the proximate and distant conditions that present boundaries to the generative potential of such digital products and their development. Considering the role of contextual conditions beyond the immediate setting of the phenomenon and understanding the interrelationships among these conditions would help in better management of opportunities and constraints shaping the evolutionary trajectories of such complex digital products.

Acknowledging digital product innovation as a process, this chapter deployed the adapted choice framework to contextualise and analyse the interdependency between the architecture, ecosystem, and environment. In order words, the role of the digital architecture and its proximate and distant context for the unfolding of digital product innovation. Through an exploration of the disaggregated interrelationship between the digital product innovation phenomenon and its overall context, it showed how the generative potential of digital product systems are bounded by inter-locking opportunities and constraints present across different layers of context and managed by the relational dynamics within its immediate setting. Bounded generativity as a notion springs from the deliberative – evolutionary process of innovation where generativity ensures that evolution is a continually unfolding process while deliberate design choices in each phase structure this trajectory in concretised operational steps to move towards the overarching aim of relevance within a practical context. Within the hybrid architecture, generativity expands the scope for deliberative experimentation while the technical interdependency between its generic and integral components ensure innovation efforts occurs within the bounds of the parameters and assumptions in the rule sets. Updating of the rules sets themselves are a result of complementary resource – relationship configurations as the ecosystem and the architecture coevolve from formation to sustenance. These resource – relationship configurations also become the way to manage the diverse and sometimes unexpected environmental drivers within the innovation process. The analysis shows that the process of digital product innovation is a function of inter-locking interdependencies between the technical and the social aspects of its digital architecture, ecosystem, and environment. Distant environmental conditions present opportunities and constraints that are harnessed and leveraged within the innovation process through the ecosystem. This also requires the management of tensions arising out of contradictory environmental conditions that shape existing resource limitations as well as opportunities and constraints presented by the hybrid architecture. Thus, contextualisation of the interdependencies between the architecture, ecosystem, and environment shows how the digital architecture and environing conditions (both proximate and distant) around the digital product innovation phenomenon shape the inherent generative potential of digital products and thereby its innovation trajectory.

7. DISCUSSION AND CONCLUSION

7.1 Overview of the thesis

The thesis began with the motivation to understand the innovation trajectory of complex digital products. It aimed to contribute to the digital product innovation literature by including complex digital products within the ambit of theorising about the phenomenon and exploring the conditions underlying complex product development. By positioning itself with the relevant literature, it identified three core constructs: digital architecture, ecosystem, and environment and interdependent relationships between them which the research sought to unpack. Positing a contextualised approach towards the study of digital product innovation, it employed the adapted choice framework to unpack the nature of conditions and relationships implicated within complex digital product innovation. Given complex product development proceeds through *ex ante* negotiations within temporary multi-actor alliances for particular design problems (Hobday, 1998; Hobday et al., 2000), digital product innovation was viewed as a process and the data collected was organised into process research comprising of phases or stages that determine the direction and nature of the innovation process. This helped to understand the development of the product, choices associated with each stage of development and the conditions underlying them. The organisation of data into process research led to descriptive accounts of the digital architecture, its proximate context of the ecosystem, and distant context of the environment by understanding cumulative outcomes of successive stages. It also helped identify the mediating conditions of resources, relationships, and structural conditions. This included shedding light on resource creation and persistence over time as well as the different environmental drivers at play and the nature of projects and stakeholders involved. This highlighted flows from multiple planes of influence and helped uncover the persistence of enduring and operational choices and trade-offs made in their wake. These descriptions and iterative engagement and re-engagement with the data and *a priori* constructs helped develop causal explanations for complex digital product innovation by understanding the interrelationships between the core constructs. These are in the form of complementary resource-relationship configurations at each successive stage with the ecosystem mediating environmental conditions, resource limitations, and technical considerations of the hybrid architecture that shape the feasibility of action. These mutually shape the generative potential of digital technologies by bounding them within opportunities and constraints created by this trilateral interdependency between digital architecture, ecosystem, and environment. The interrelationships between them helped arrive at the concept of bounded generativity as a contextualised approach and an explanatory construct for the deliberative-evolutionary trajectory of complex digital product innovation. The following sections highlight distinction of Flee ABM as a complex digital product and the concept of bounded generativity in the light of existing literature on digital product innovation and how it extends or complements it. Then it goes on to discuss the contextualised approach to the study of digital product innovation, implications for practice, contributions and limitations, and pathways for future research.

7.2 Flee as a complex digital product

The hybrid architecture of Flee exhibits characteristics of a complex product. It is comprised of multiple interconnected elements many of which are customised to suit the needs of both the model and application area (Hobday, 1998; Hobday et al., 2000). As a complex product Flee continues to evolve with changes in underlying technologies, rule sets, parameters and assumptions, computational capacity or software development approaches. The resource – relationships configurations at each stage showed the sheer diversity of specialised expertise not just within the team but also within the successive alliances with ecosystem participants that bring together complementary resources, often creating newer ones leading to antecedent conditions for product evolution in successive phases. However, as a complex *digital* product Flee exhibits a hybrid architecture comprising generic software components like FabFlee, MUSCLE 3, and VVUQ toolkits which are tightly coupled with the main code through rules sets. While these modular digital components enable functional extension and experimentation such changes must occur within the bounds of its integral composition. Further, while approaches like the SDA makes them adaptable across contexts, this adaptability is structured by existing functionalities currently allowed by the code. Extensions of functionalities like coupling or multiscale modelling require further customisations. Each of these extensions are brought about by resource – relationship configuration that results in the *ex ant*e negotiation of the design problem by mediating extant environmental conditions and technical considerations of the digital architecture. This is in contrast with existing digital product innovation literature in which product owners negotiate tensions within the innovation process *ex post* or after the fact when particular behaviours are observed among ecosystem participants (Ghazawneh & Henfridsson, 2013; Eaton et al., 2015).

Much of this has to do with the nature of the architecture and how the tasks get distributed. Within layered modular architectures, the standardisation achieved through modularity distributes design tasks across a wide range of participants and helps leverage bi-directional complementarity through scale of innovation and low marginal cost of combining digital complements (Yoo et al., 2010; Jacobides et al., 2018; Faulkner & Runde, 2019). Hybrid architectures within complex digital products like Flee have to negotiate the tension not just between its generic and integral components but with the opportunities and constraints across layers of environing conditions. Within complex digital products, the tension arises because the vast complexity of social reality and environmental conditions refuse to be tamed through simplified causal relationships and requires multifaceted and multidirectional search and learning (Clark, 1985; Hobday, 1998; Hobday et al., 2000). These are then managed through a stable evolvability that ensures successive resource – relationship configurations. The condition for stable evolvability is required for the process to accommodate successive *ex ante* negotiations for arriving at operational choices and partial resolutions of attendant tensions within the innovation process.

Stable evolvability is a condition whereby the financial resources provided by funded projects offered sustainability for the team to maintain the product development process. Existing digital product innovation takes mass market conditions and the firm environment as a given as seen in paradigmatic examples of mobile operating systems and app stores (Yoo et al., 2010; Ghazawneh & Henfridsson, 2013; Eaton et al., 2015). However, the development of complex products has to negotiate the key financial element of creating conditions for resources and relationships to align through an opportunity structure. While digital components or complements and open source environment enable low marginal cost of combination (Faulkner & Runde, 2009), this can only be realised when the opportunity structure brings together other resources including financial resources into conducive configurations.

Digital products like Flee involve manipulation and leveraging of the conditions of digitality to converge on reality and analytically render it through code (Kallinikos, 2009). In doing so, it approximates not just its problem context, i.e. refugee movement during forced displacement, but transmutes a host of other environmental conditions like open science policies, socio-political demand conditions, data limitations, and ethical considerations into product development. These conditions of possibility across layers of context then come to be implicated within the product. Complex products are usually developed for high reliability industries like airlines where product failures can have devastating consequences (Ulrich, 1995; Hobday, 1998; Hobday et al., 2000). The humanitarian sector is fraught with protection concerns for some of the most vulnerable people in the world who have escaped war, persecution, and natural disasters that have wiped away their lives and livelihoods. The failure of predictive models in meeting humanitarian operational needs in crisis situations can determine resultant resource allocations with the potential to exacerbate already precarious situations. These require additional considerations for the way innovation is organised and managed within the sector. It includes attendant issues like inequalities in funding across locations that can constrain the model development process for a particular conflict, which in turn are affected by geopolitics which can draw attention away from an existing conflict to a more rapidly unfolding or strategic one. This places additional responsibilities to ensure the assumptions on which the model is predicated is related as closely as possible to the conditions on the ground but avoid the pitfalls of overfitting to ensure applicability across contexts. The development of complex products like Flee exhibit high user involvement by humanitarian organisations where such involvement ranges from tacit support and knowledge exchange, to direct requirements specifications, and sharing of operational knowledge. The ad hoc collaborations with humanitarian organisation also meant working under conditions of uncertainty in terms of requirements specifications (Clark, 1985; Hobday, 1998, Hobday et al., 2000). This further reinforced the need for stable evolvability enabled by multiple funded projects which could be harnessed and translated into product development that remains relevant for the humanitarian context as opposed to just being within the remit of scientific computational advancement.

Developing the products under such uncertainty and moving from one resource – relationship configuration to another were characterised by partial resolutions and incentives beyond limited economic ones. For example, collaborations with humanitarian organisations did not involve any financial exchange but was based on reciprocal utility. Information resources provided by humanitarian organisations helped the team take the model closer to reality. Such collaborations in turn helped humanitarian organisations in co-developing a workable predictive computational model. This ties into the sector level impetus on policy and practice in moving towards data driven decision-making and predictive analytics. This insight was reinforced through the KEI workshop 2022 which highlighted that achieving the technical-organisational-contextual fit depended on a wider range of considerations that different stakeholders bring to bear as well as limitations around attendant organisational capacities and fragmentations within the sector.

Flee demonstrates how the complex nature of the product and its hybrid architecture shapes and structures the way innovation in organised. The hybrid architecture of Flee requires complex and diverse specialisations to work together to bring about change and product evolution. However, this can only happen through successive multi-actor alliances focused on a particular design problem. This is because while the generic modular aspects enable participation and experimentation, this can only occur within the integral ambit of the main code. Unlike the significant power exerted by the product owner in modular ecosystems, the system integrator works at once within uncertain requirements as well as within well-defined project mandates which are negotiated at each stage through resource – relationship configurations. This is to ensure their convergence around the diverse range of influences exerted from multiple planes within a given context i.e. the proximate context of the ecosystem (in the form of professional networks and consortium partners) and across layers of context i.e. ecosystem and distant environmental context of multiple environmental drivers, and their interdependency with the hybrid digital architecture. These come to be implicated within the innovation process structuring the conditions of ongoing product development.

7.3 Bounded generativity as a contextualised approach: The deliberative-evolutionary nature of complex digital product innovation

Bounded generativity offers an explanatory construct to unpack the nature of relationships under consideration between the digital architecture, ecosystem, and environment or how the interdependencies between the architecture within and across layers of context shape the innovation trajectory of complex digital products. The notion of bounded generativity helps understand how the generative potential of digital artefacts come to depend on direct and derived conditions that arise within the process of innovation, thereby making it both a deliberative and evolutionary process. Taking a ‘product view’ of generativity, the contextualised approach to bounded generativity suggests that digital product innovation comes to be shaped both by current and antecedent conditions that translate generative potential of digital artefacts in particular ways. It acknowledges multiple planes of influence implicated within the given direction of product development rather than uni-directional relationships between the user base and product and ecosystem boundaries respectively as explored by Fürstenau et al. (2023). Moreover, complex product innovation involves institutional users like humanitarian organisation precluding the condition of user base growth in the absence of mass market conditions. Extending understandings of digitality (Yoo et al., 2010; Faulkner & Runde, 2009; 2019), it suggests cost of recombinable innovation is only low when other resources in the portfolio of are maintained through stable evolvability. The foundational phases which involved development of first version of Flee, SDA, and FabFlee led to the creation of knowledge and computational resources that became a part of successive opportunity structures leading to evolutionary conditions on which further development of Flee could be predicated. However, as a process with circular reciprocal relationships between action and outcome, each phase was also a result of deliberation and bringing together particular resource – relationship configurations. For example, successive iterations of rule sets involved moving from intuitive assumptions based on secondary research to being informed by INGOs of on-ground practical conditions. Similarly for the development of coupled approaches involving weather conditions included insights and deliberation with different stakeholders like ECMWF and academic collaborations like the University of Utrecht. The latter also led to thinking about including geospatial route planning algorithms based on terrain accessibility.

Given the uncertainty of requirement specifications, as opposed to layered modular architectures, within which such development occurs, the stable evolvability of existing resource – relationships configurations help mobilise them to respond to demand conditions like flashpoint events like the outbreak of war. Therefore, while deliberations underlying particular developments ensures the conditions of evolution, the evolving and dynamic nature of digital product development also creates conditions for new deliberative development throwing up new challenges and design problems when requirements become clearer or certain resources become more available. These new and antecedent conditions become the way the innovation progresses through trade-offs or partial resolutions of attendant tensions, the nature of relationships within the proximate context, and the strength of resource – relationship configurations to mediate given environmental drivers that becomes prominent for particular design problems during given phases. Therefore, bounded generativity as a contextualised approach is at once how complex digital product innovation progresses and the way the innovation process itself is organised.

Extant literature on digital product innovation has been focused on understanding evolutionary dynamics of digital products depending on the extent to which the product owner is able to manage or negotiate external economic environmental conditions like multihoming and switching costs and third party behaviour anchored through modularity (Tiwana et al., 2010). The notion of bounded generativity opens up ways of thinking about digital product innovation that includes a wider set of conditions beyond economic and managerial ones. It also stands in contrast with the implicit assumptions within digital product innovation literature of unbounded generative potential of digital technologies in driving the innovation process as a function of their specific digital architectures (Yoo et al., 2010; Kallinikos et al., 2013). This helps think of the ways generativity of hybrid architectures of algorithmic systems like ABM acts and is acted upon by extant social reality and the way they come to shape each other in an increasingly computationally intensive world. It suggests that theorisation around digital product innovation needs to consider how diverse environmental conditions need to be managed. It also highlights the ways such conditions are managed by leveraging complementary resources and relationships within the extant ecosystem and proceed through partial resolutions. This highlights how layered and relational conditions come to shape and be shaped by composition of the digital architecture. These inter-locking interdependencies helps appreciate operational choices in light of enduring ones to acknowledge the diversity of conditions that are implicated within the digital product innovation process. It proposes a rethinking of conditions for ecosystem emergence (Jacobides et al., 2018) and how digital architectural forms shapes mode of organising and managing its innovation (Ghazawneh & Henfridsson, 2013) and the tensions therein (Wareham et al., 2014). It highlights that modularity is not the only pre-condition for ecosystem emergence. A hybrid architecture organises its innovation ecosystem into successive resource-relationship configurations through task definition and specialisations which help it negotiate successive design problems through partial resolution. Further, while existing literature contains some references to the role of wider environmental conditions, they had not been previously theoretically included. As an illustrative example, existing studies on digital product innovation have not considered the role of changing policy environments, legal requirements, and how legal interpretations of the output across jurisdictions can work at cross purposes with it intended objectives. Transnational policies around responsibility and safety affect even mature social media products requiring changes to how their systems operate. Apart from providing a holistic and contextualised approach to thinking about digital product innovation, this study helped suggest a way to think about complex digital products and emerging technologies and the conditions that need to be taken into account for their management. Emerging technologies like large language models like ChatGPT exhibit a long investment horizon with significant capital investment with limitations structured by representativeness of the dataset and considerations around responsible practices with regard to their collection (Wiggers, 2023; Kumar, 2023). These present significant innovation challenges in developing such technologies in a responsible manner that instils trust in their use. While ChatGPT might have been made amenable for recombinant innovative use, the underlying product remains an integral composition of multiple technologies, data, computational and knowledge resources. This is similar for other emerging technologies finding diversifying applications across domains of health, education, socio-economic development, and humanitarian and disaster management. Such safety critical fields require a more holistic attention towards understanding the dynamics of their innovation processes. This is because the structure of trilateral interdependency will have significant implications for the nature of innovation tensions and the processes whereby they come to be negotiated and managed.

7.4 Adapted choice framework: A contextualising approach for digital product innovation

This study took a contextualised socio-technical approach to expand the scope of theorisation within digital product innovation. A contextualised socio-technical approach was required to develop holistic explanations and avoid the under-socialised or over-socialised dichotomy in the treatment of IS phenomenon (Avgerou, 2019; Rush et al., 2021). It helped focus attention not only on the specificity of the digital architecture but how diverse environing conditions stand in inter-locking relationships with it. However, contextualising a digital IS phenomenon required an understanding of how generativity has shaped existing modes of organising for innovation across and within layers of contexts. It also involved negotiating enduring tensions within contextual research of scale, scope, and detail (Avgerou, 2019). The adapted choice framework helped provide a contextualising approach to manage these extant issues within an analytical framework. The framework helps link actions, outcomes, and conditions. Successively applying it across key phases enabled the mapping of granular aspects of the digital architecture, ecosystem participants, and environmental conditions to unpack resources, relationships, and structures that produce observed outcomes. The cumulative overview of the process helped identify broader patterns and interdependencies that form the basis of explanations for causal trajectories driving the digital product innovation process. In conjunction with process research, its resolution of scope, scale, and details lies in its disaggregated focus on technology, resources, relationships, and structure which provide the basis for tracing upward causation (Winter et al., 2014) through successive and cumulative outcomes. This study proceeded from a recognition of a trilateral interdependency between digital architecture, ecosystem, and environment which it sought to unpack. The notion of bounded generativity helps explaining the inter-locking conditions underlying such interdependency which structures and bounds the generative potential of complex digital products. The adapted choice framework provides an approach for future contextualised studies of digital product innovation by taking into account both social and technical aspects and conditions beyond the immediate setting of the environment. It provides a holistic and process oriented approach to theorising about IS phenomenon to accommodate the multiple planes of influence and conditions that are implicated within it (Fichman et al, 2014; Pettigrew, 1997). The development and application of the adapted choice framework towards a contextualised study of digital product innovation submits a theoretical approach in answer to a question raised by Avgerou (2019): Given the predominant view of context as a social domain and a theoretical view of IS as a socio-technical phenomenon, what would a socio-technical contextual IS theory look like?

7.5 Implications for practice

Emerging technologies not only alter the productisation process but also the whole end-to-end development structures calling into question sustained profitability and sustainability of the entity developing the product (Hamid & Suoheimo, 2023). The magnitude of digital transformation in society and economy by a variety of new and emerging technology in terms of computing power, analytical sophistication, and bandwidth is reshaping how digital product innovation occurs. This has led to a number of interrelated shifts where technology acts as a centrifugal force pushing innovation to expert networks and the need for radical new approaches and continuous learning built around skills required to solve emerging design problems. This means that power is no longer centralised at a system integration level with data intensive technologies creating more downstream risks and responsibilities in terms of security and data protection (Van Kuiken, 2022). Addressing the persistent gap between technology-contextual fit at the point of implementation, key managerial challenges include the variety of internal markets to be served, resistance to change, choice of implementation site, and someone to take overall responsibility (Leonard-Barton & Kraus, 1985). In the light of 'myriad, inter-related deep technologies that span different stages of evolution' a host of other conditions come into play which includes funds flows, intellection property investments, developing existing technological capabilities and demand conditions (BCG & NASSCOM, 2022) which requires an in-depth understanding of both the technology and its proximate and distant contexts.

This study submits both theoretical insight and a practical analytical framework for managers and developers to make contextualised decisions within their domain of application and work towards removing incongruences in arriving at a technological-organisational-contextual fit. This becomes particularly important when leveraging existing solutions, forging the required expert coalitions, and arriving at appropriate resource-relationship configurations to drive the innovation process. Unpacking the diverse conditions underlying successive design choices also helps manage downstream risks and responsibilities to understand the eventual social implication associated within particular trade-offs and design choices. This is because these choices can have far reaching consequences at implementation given concerns around discrimination and exclusion associated with emerging technologies (Markus & Nan, 2020; Kallinikos & Hasselbladh, 2009). In safety critical contexts like the humanitarian sector, they would come to influence strategic decision-making for relief and rescue operations, aid delivery, resource allocation and provision of basic services to some of the most vulnerable populations in the world. This becomes particularly important as problems within the humanitarian system continue to grow in scale, scope, and complexity combined with continued pressure on available resources (Rush et al., 2021). Particularly as digital technologies are developed away from the local contexts whose operations they seek to influence. This exposes situated agents to messages, information, and requests arriving from distant events or sources, thereby undermining initiatives aimed at longer-term resilience and local capacity building. These disembedding qualities succeed in establishing new considerations of action and control that require serious consideration (Kallinikos, 2006), Every technological system has a set of embedded norms (Lessig, 2000) that come into play on usage (Wajcman, 2004). Many of the social implications highlighted are structured through the design and development process which come into force much before the technology is implemented and essentially circumscribes the eventual human-technology interface. While social implications of digital technologies have been studied within IS (Majchrzak et al., 2016), there is still a need to understand the processes by which these changes occur and that they might be a consequence of the distinctive characteristics of digital technologies themselves (Markus & Nan, 2020). Just as the humanitarian sector presents particular concerns, so might other fields of application. The advent of large language models and generative artificial intelligence has propelled policy development on account of potential harms, exclusion, lack of representation, transparency and accountability which need to be factored into the process of design and development to prevent ex post radical changes (Djeffal, 2023). Given the diversity of technologies dominating industry insights and technology trends, it becomes imperative within a given field of application to understand how trilateral socio-technical interdependencies are structured within a given domain.

7.6 Flee and the broader humanitarian technology landscape

As mentioned earlier in the study, Flee was developed within the wider environmental context of a policy push towards anticipatory action within humanitarian management. Anticipatory action involves the use of predictive technologies to pre-position resources with the aim to mitigate adverse consequences of conflicts and sudden onset disasters. Predictions and forecasts can help flag humanitarian needs arising from different shocks or crises which opens a window of opportunity for humanitarian actors to reduce the overall impact by acting before such needs materialise (UNOCHA, n.d.). New and emerging technologies with predictive and forecasting capabilities are increasingly being centred within humanitarian policy to support this paradigm shift by enabling faster and more effective humanitarian action (UNOCHA, 2021). Currently, a wide variety of computational approaches and predictive models exist at varying levels of maturity for different humanitarian crises. These range from predictions for the level of severe flooding to the duration and intensity of conflict related displacement (Jelonek, 2023). In parallel, a diversity of computational approaches exist ranging from artificial intelligence and machine learning models to unmanned aerial vehicles or drones, Internet of Things, and satellite imagery and remote sensing tools (UNOCHA, 2021). Apart from the ABM approach discussed in this study, different computational approaches are developed and applied for different crisis scenarios depending on their individual functionalities. Artificial intelligence and machine learning enable analysis and interpretation of vast and complex datasets to improve predictions and inform decision-making. Within the migration context, apart from the ABM approach discussed in this study two key machine learning approaches currently exist, UNHCR's Project Jetson and the

Danish Refugee Council's Foresight model. Both models aim to provide a macro-level overview of migration trends with multiple indicators and parameters with the latter incorporating 120 such indicators. The Foresight model currently covers 26 countries and is said to perform 'quite well' with more than half the forecasts produced being less than 10% off from actual displacement figures (DRC, n.d.). While the Foresight model is currently in use, Project Jetson was an experimental approach launched by UNHCR in 2017. It was modelled on Somalia and was intended to provide a use-case example for predicting forced displacement. Experimentations with Jetson led to pilot projects like the SatCropper predictive model which aims to find the nexus between climate change and forced displacement (Project Jetson, n.d.; UNHCR, n.d.). While ABM, artificial intelligence and machine learning enable interpretation of data to generate actionable insights the others provide sources of data collection (UNOCHA, 2021). Unmanned aerial vehicles or drones and remote sensing technologies can speed up post-disaster risk assessment, mapping, and monitoring. They are cheaper than satellites and are flown autonomously on a programmed route to collect imagery and enable faster and more efficient response planning and monitoring (American Red Cross & IFRC, 2015; UNOCHA, 2021). However, unlike satellites, drones depend on close proximity to pilots. The World Food Programme used drone imagery and machine learning to improve post-disaster damage assessment which needs to be conducted within 72h after a disaster. Drone images were run through image classification algorithms to map and quantify the number of houses damaged in the aftermath of a cyclone (Codastefano, 2019). Internet of Things as networks of interconnected devices help transmit real-time data which enhances the capacity for network analytics, thereby improving early warning and operational efficiencies. Internet of Things are increasingly becoming important for humanitarian logistics and have been used for cold chain management of vaccines and for monitoring and managing temperature controlled supply chains (UNOCHA, 2021). Further, remote sensing satellite imagery have been used for anticipatory action in climate induced humanitarian crisis (Whelan & Verity, 2022). In Bangladesh, flood forecasting was made possible due to a combination of machine learning, satellite imagery, spatial mapping (UNOCHA, 2021). Unmanned aerial vehicles, Internet of Things, and remote sensing satellite represent data collection and transmission technologies which then have to undergo further analysis to generate actionable insights for decision-making. Data from these sources are often combined with other data points and modes of analysis like artificial intelligence and machine learning.

Some of the issues emanating from the use of such data from diverse sources include the lack of open, accurate, and timely data along with unstructured data with concerns around data protection risks. This is exacerbated by the lack of data protection regulations and country specific regulations on drones. Moreover, the use of drones for conflicts makes them a sensitive mode of data collection for humanitarian management of forced displacement posing privacy and security challenges for those fleeing violent conflict. In a conflict zone, drones can further interfere with manned aircraft conducting search and rescue operations and supply drops (Joseph et al., 2020). Satellite imagery can be expensive with poor resolution, impacted by atmospheric conditions like cloud cover, and is not available in a timely manner (Joseph et al., 2020). Internet of Things can be subject to cyberattacks and the hardware can be cost-prohibitive requiring regular maintenance with the current state of technology hindering rapid deployment. It also amplifies data protection and privacy concerns because of the potential to transmit personal and sensitive information (UNOCHA, 2021). These considerations limit the data available for uptake in computational modelling techniques like ABM or machine learning.

While ABM models are constructed based on underlying assumptions about modes of relationship and behaviour between the agent and the environment, machine learning models tend to proceed backwards from the end goal i.e. the quantity to be predicted with the choice of model intended to minimise a given error metric. Once prediction targets and error functions are chosen the process can be agnostic to the choice of predictive features and as a result, they rely less on theoretical assumptions about migration (Pham & Luengo Oroz, 2022). In other words, machine learning models thrive on large numbers of parameters within the model while ABM models explore the detailed relationship between a few parameters and the phenomenon. However, one of the criticisms of machine learning models is their ‘black boxed’ nature or the lack of interpretability (Pham & Luengo Oroz, 2022). Therefore, it may become necessary to conduct additional analysis or select related explainable algorithms at the design stage to identify and understand how different features influence the resulting forecasts. ABMs are particularly helpful for scenario-planning under conditions of data paucity to understand different patterns of an agents’ actions in different scenarios and thereby help arrive at the nature of different outcomes under different conditions and assumptions. ABMs help in developing insights into system dynamics as opposed to optimising predictive accuracy of the models as in machine learning (Pham & Luengo Oroz, 2022).

However, the notion of accuracy itself can be problematised. While in mathematical terms it refers to optimisation of error rates, such validation exercises to calculate error rates remain fraught on account of the revision of UNHCR numbers as a result of overcounting and undercounting and the methodological drift that some of the major datasets suffer in the sector (KEI Event 2022). Since ABMs tend to be driven by scenarios and insights into the phenomenon, multiple causal factors were identified either through secondary research or on the basis of insights shared by humanitarian organisations and NGOs on the ground where optimisation became an iterative exercise to computationally codify a dynamic social reality. While social reality involves a multiplicity of factors inflecting the phenomenon at a given point in time, including all of them within the model would make the model computationally expensive. It would also lead to overfitting to a particular context thereby rendering it ineffective to others. Moreover, it becomes difficult to fully optimise because displaced persons tend to take decisions based on tacit knowledge while on the move such as gravitating towards camps of same ethnicity or knowing that conflicts shift elsewhere when certain regions become impassable on account of seasonality for which there is no data to validate against.

Since up until the point of observation, Flee had not been used for on ground implementation, it was not possible to know how accurately it predicts currently unfolding conflicts. The deployment of predictive technologies in anticipatory action proceeds through pre-agreed frameworks. The UN OCHA facilitates collective anticipatory action and coordination of response driven by three core elements based on (1) a forecast based trigger elicited by the computational model, this leads to (2) release of pre-arranged finance and (3) mobilisation of an implementation plan that is pre-agreed among multisectoral UN agencies, INGOs, and local NGOs (UNOCHA, n.d.). There is currently no sector wide standard for defining a threshold for a trigger, they vary with computational technique and crisis scenario. For e.g. the anticipatory action pilot that used a machine learning model to determine the relationship between wind speed and housing damage in the Philippines to mitigate the impact of typhoons had a two-stage trigger activation: (1) readiness trigger (pre-activation) with 4-7 days lead time for predicting a tropical cyclone with the potential to reach Category 3 level or higher i.e. greater than 178 km/hr maximum 1-minute sustained wind speed or 158 km/ hr maximum 10-minute sustained wind speed. (2) The activation trigger to be initiated on or before 72 hours prior to projected landfall wherein an impact map is produced based on the forecasts of the predicted number of totally damaged buildings. The action plan is activated if the number of totally damaged houses fall within the range of 50% probability that 80,000 houses will be totally damaged or 50% probability that at least 5000 houses will be totally damaged (UNOCHA, 2022). In December 2021 Typhoon Rai (locally known as Odette) lashed through southern and central Philippines rapidly intensifying within a matter of a few hours causing damage and destruction to 2.1 million houses and 10.2 million hectares of agricultural land making it one of the deadliest and costliest typhoons to have hit Philippines. The forecasting model was unable to pick up the rapid intensification in the last hours before landfall; this could potentially be attributed to affected regions within the country falling outside the pilot areas for which the model was being tested. The after-action review of the incident highlighted the importance of further studies on rapid intensification and incorporation of more areas under the pilot programme. The Philippines is one of the more well documented anticipatory action frameworks (UNOCHA, 2022). Documentation of existing frameworks for other countries show the combined use of existing global and national forecasts like in the case of flood forecasting for Bangladesh with a 15 day lead time for 50% probability of an adverse event and 5 day lead time for breach of a government defined ‘danger level’ for readiness and activation triggers respectively. (UNOCHA, 2023).

The above discussion highlights how the model development and implementation processes are imbued with successive value choices embedded in design of both the computational approach as well as the implementation plan (Baharmand et al., 2021; Read et al., 2021). This ranges from simple statistical correlations to represent significant relationships between variables to qualify for anticipatory action or the choice of triggers and threshold levels which real events force developers and implementors to recalibrate and optimise. These choices embedded in efficiency generating processes powered by mathematical and statistical formulations and algorithms used in predictive analytics shape how needs and knowledge are represented that comes to determine the exclusions and the distribution of benefits and burdens (Burns, 2018) e.g. in the Philippines case the anticipatory action plan was stood down until rapid intensification since the typhoon had not met any of the threshold criteria (UNOCHA, 2022). Consequently, they increase the risk of failure in the humanitarian context due to functional defects or human error at the point of implementation (Sandvik et al., 2014). This is compounded by the fact that predictive technologies are heavily reliant on data which is extremely fragmented within the humanitarian sector and this limited and patchy data comes to be imbued with knowledge which shapes assumptions and selection of parameters which structures the visibility and inclusion of affected populations within the modelling process (Jacobsen & Fast, 2019). As such, decision-making based on such outputs blurs the line between care and control with implications for bias, discrimination, and procedural fairness (Jacobsen & Fast, 2019; Molnar, 2020; Molnar & Gill, 2018). Moreover, as technologies come to be developed away from local contexts by experts and professionals it marks a move towards increased centralisation of decision-making away from the field leading to remote management of crisis through iterative determination of what counts as optimal thresholds for intervention (Read et al., 2016). As predictive technologies receive increasing thrust within operational planning, they are coming to represent the 'digital recoupment of the consequent loss of face-to-face contact' (Duffield, 2015). The capacity of such models for monitoring, sorting, and classification heightens the potential of such models becoming instruments of control regulating local outcomes (Gandy, 2021). This is because their implementation often takes place in fragile contexts with limited regulatory and governance oversight where state accountability is weak (Molnar, 2020). Further, the absence of international regulations increases the risk of violation and infringement of human rights, thereby heightening risks for already vulnerable populations by exposing them to adverse and unintended consequences arising from the use of digital technologies.

7.7 Contributions and limitations

Research is an evolving enquiry where *a priori* constructs form the first step in a long series of gradual precisions (Lund, 2014). The methodology chapter highlighted the pathway for arriving at theoretical contributions by asking the question 'of what is this a case?'. The answer to this question involves moving across an analytical matrix from the specific and concrete to the general and abstract. The research identified specific and concrete aspects of the case under consideration in the form of choices associated with different phases in the process of digital product innovation and descriptions for the architecture, ecosystem, and environment and resource categories, relationships, and structural conditions that emerged from process data. These specific and concrete observations helped identify the nature of relationships and interdependencies which in turn led to the concept of bounded generativity i.e. complex digital product innovation as a deliberative – evolutionary process whereby the generative potential of digital products is shaped and structured by the direct and derived conditions. These include both antecedent conditions in the form of persistence of resources created in earlier phases and successive complementary resource – relationship configurations for particular design problems. Using specific and concrete observations and descriptions as first order constructs, the study arrives at the notion of bounded generativity as a second order construct or a portable concept that provides insight and highlights theoretical implications for future study (Van Maanen, 1983; Lee & Baskerville, 2003; Falleti & Lynch, 2009). This helps present theoretical propositions that through further research can be consolidated into theory. Acknowledging that theory development at the level of the disciplinary field is incremental, this research submits the theoretical proposition of digital product innovation that is shaped by how the inherent generative potential of digital technologies in bound by the nature of trilateral interdependency and associated direct and derived conditions over time. This proposition is submitted as contribution to the field of digital product innovation in a way that provides insight on the conditions underlying innovation for alternative architectural forms beyond the ones that have been studied in the literature so far. It also provides a holistic approach towards understanding the phenomenon of digital product innovation. Generalisation through development of portable concepts and putting forward theoretical propositions, particularly from contextualised studies is to enter into a dialogue to understand how this research resonates with other works (Lund, 2014). The development of theoretical propositions does not lay claim to universal validity or establish actual validity but suggests likelihood and probability. Submitting these provisional theoretical propositions through research are a form of scholarly communication where future work will confirm or contradict the proposition's generalisability.

While this study is contextualised within its particular domain of enquiry, it provides a flexible explanatory construct in the form of bounded generativity and a framework of analysis as a process for unpacking contextual socio-technical conditions at the level of the technology, ecosystem, and environment. Thereby suggesting not just a need for a contextualised approach to digital product innovation but a framework of contextualising similar studies in the different domains of application. Bounded generativity as a contextualised approach is offered as a portable concept that is derived from the specificity of this particular case but without specifying outcomes in and for studies in other contexts thereby enabling transferability for future studies. Consequently, bounded generativity as an explanatory construct does not have predictive value and as causal explanation in qualitative research is indeterminate in its proposition (Avgerou, 2013). This is because cases are essentially an edited chunk of an empirical reality where only certain aspects are observable and taken into consideration (Lund, 2014; Townley, 2008).

This research was based on a single case study which helps in studying a particular case in the full richness of it detail. However, the degree of generalisation from contextualised case studies improves incrementally through comparative case studies (Avgerou, 2019; Pettigrew, 1997). In comparative research, domains of enquiry which are considered the source of contextual influence become an important determinant of similarity or difference that needs to be made explicit in order to allow for the comparison of research findings. Thus, theory building becomes a continuous work in progress both within the researcher's own workstream and among the community of researchers in a disciplinary area within which the given contribution resonates. Therefore, the proof of the contribution remains in its utility (Lund, 2014) i.e. whether it can help other researchers in other contextual settings understand their work as much as it has helped the given researcher in understanding and explaining the phenomenon under consideration.

7.8 Conclusion and way forward

Digital product innovation is a comparatively understudied phenomenon with limited focus on explaining how systems evolve over time and limited attention given to incorporating the nature of architecture within theory development (Yoo et al., 2010; Orlikowski & Iacono, 2001). This study has been an attempt like Yoo et al. (2010) to bring architecture back into theorisation and in doing so expand the nature of architecture under consideration within the discipline. It has worked to develop holistic explanations to build an IS theory and approach that takes into account the multiplicity of conditions to move beyond perspective-centred theorisation borrowed from other fields (Fichman et al., 2014). As mentioned above, every contribution is a form of scholarly communication and this study forms the first step in moving towards a general theory on complex digital products and complex digital product innovation. While this study focused on a particular type of architecture i.e. an ABM model, it remains to be seen how other forms of emerging technologies like different types of machine learning models, large language models, blockchain, deep learning, convolutional neural networks among others shape their innovation processes. Wherein, it must be considered that some of this innovation might reflect the dark side of bounded generativity in the form of deep fakes and misinformation, racial bias in facial recognition systems, or gender and racial bias in commercial artificial intelligence products among others (Satariano & Mozur, 2023; Najibi, 2020; Buolamwini & Gebru, 2018; Raji & Buolamwini, 2019).

It highlights significant tensions in the innovation process and the impact of environmental drivers like the representativeness of datasets used to train models and the socio-political and political economic histories behind their make-up, such as the tension between ensuring representativeness through data collection while maintaining privacy, and the composition of the multi-actor expert coalitions in the proximate context in terms of who gets to make design decisions and about whom (Tucker, 2017; private conversation; Costanza-Chock, 2020). In sum, the future directions of this research would not only have implications for how complex product innovation can be better managed but also how they can be more responsibly managed. This involves acknowledging that the scope of digital product innovation stretches not only through positive digitally mediated social and economic transformations but also the underexplored dark underside of how digital product innovation with negative social implications are enabled through conditions that in many ways are similar to downstream value added recombinant innovation (Zittrain, 2006; 2008).

In more managerial focused research, it provides the entry-point to understand the governance of complex digital product ecosystem and the orchestration of activities therein with the system integrator as the unit of analysis. The nature of trilateral interdependency as well the constituent components of each element would help to explore criteria for strategic decision-making in relation to product development. In theory focused research, it remains to be seen how material objectification and computational rendition of reality into code within emerging technologies likes the ones mentioned above shapes material and social agency through materiality of technology and the materialisation of social reality through analytical reduction. It helps move towards developing a theory of complex digital products based on comparative studies. The ability to do an in-depth research as a result of the access provided by the BUL team enabled deep engagement with the Flee project which helped develop the criteria, characteristics, and vocabulary to undertake such future research. It helped provide a flexible framework and analytical categories to understand the conditions of the phenomenon under consideration and serve as building blocks for further research.

8. BIBLIOGRPAHY

Abar, S., Theodoropoulos, G. K., Lemarinier, P., & O'Hare, G. M. (2017). Agent based modelling and simulation tools: A review of the state-of-art software. *Computer Science Review, 24*, 13–33. https://doi.org/10.1016/j.cosrev.2017.03.001.

Abernathy, W.J. & Utterback, J.M. (1978). Patterns of industrial innovation. *Technology Review, 80*(7), 2–9.

Adner, R. (2017). Ecosystem as structure: an actionable construct for strategy. *Journal of Management, 43*, 39–58. https://doi.org/10.1177/0149206316678451.

Akter, S., Motamarri, S., Sajib, S., Bandara, R. J., Tarba, S., & Vrontis, D. (2021). Theorising the Microfoundations of analytics empowerment capability for humanitarian service systems. *Annals of Operation Research.* https://doi.org/10.1007/s10479-021-04386-5.

Alaimo, C. & Kallinikos, J. (2017). Computing the everyday: Social media as data platforms. *The Information Society, 33*(4), 175–191. https://doi.org.10.1080/01972243.2017.1318327.

Alaimo, C. & Kallinikos, J. (2019). Recommender systems. In: T. Beyes, R. Holt, & C. Pias (Eds.). *The Oxford Handbook of Media, Technology and Organization Studies* (401-411). Oxford University Press.

Alaimo, C., Kallinikos, J., & Valderrama, E. (2020). Platforms as service ecosystems: Lessons from social media. *Journal of Information Technology, 35*(1), 25 - 48. https://doi.org/10.1177/0268396219881462.

Alexander, C. (1964). *Notes on the synthesis of form.* Harvard University Press.

Allan, R. (2010). *Survey of agent-based modelling and simulation tools*. Science and Technology Facilities Council. Available at: http://purl.org/net/epubs/manifestation/5601/DLTR-2010-007.pdf.

Alsop, R. & Heinsohn, N. (2005). Measuring empowerment in practice—structuring analysis and framing indicators. *World Bank*. https://doi.org/10.1596/1813-9450-3510.

American Red Cross & IFRC. (2015). *A vision for the humanitarian use of emerging technology for emerging needs*. Retrieved from https://preparecenter.org/resource/a-vision-for-the-humanitarian-use-of-emerging-technology-for-emerging-needs/ [07 April 2024]

Anderson, J., Chaturvedi, A. & Cibulskis, M. (2007). Simulation tools for developing policies for complex systems: Modeling the health and safety of refugee communities.

Health Care Management Science, *10*(4), 331–339. https://doi.org/10.1007/s10729-007-9030-y.

Anderson, J., Chaturvedi, A., Lengacher, D. & Cibulskis, M. (2006). Modeling the health of refugee camps: An agent-based computational approach. *Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems*. https://doi.org/10.1109/CBMS.2006.118.

Arora, P., Malhotra, P., Lakhmani, I. & Srivastava, J. (2017). Agent based software development. *International Journal of Scientific Research Engineering and Technology*, *6*(4).

Avgerou, C. (2000), Recognising alternative rationalities in the deployment of information systems. The Electronic Journal of Information Systems in Developing Countries, 3(1), 1-15. https://doi.org/10.1002/j.1681-4835.2000.tb00021.x.

Avgerou, C. (2001), The significance of context in information systems and organizational change. *Information Systems Journal*, *11*(1), 43-63. https://doi.org/10.1046/j.1365-2575.2001.00095.x.

Avgerou, C. (2013). Social mechanisms for causal explanation in social theory based IS research. *Journal of Association for Information Systems*, *14*(8), 399-419. https://doi.org/10.17705/1jais.00341.

Avgerou, C. (2019). Contextual explanation: Alternative approaches and persistent challenges. *MIS Quarterly*, *43*(3), 977-1006. https://doi.org/10.25300/MISQ/2019/13990.

Avgerou, C. & Madon, M. (2004). Framing information systems studies: Understanding the social context of IS innovation. In C. Avgerou, C. Ciborra, & F. Land. (Eds.). *The Social Study of Information Technology* (162-182). Oxford University Press.

Avgerou, C. & McGrath, K. (2007). Power, rationality, and the art of living through socio-technical change. *MIS Quarterly, 31*(2). https://doi.org/10.2307/25148792.

Baharmand, H., Saeed, N., Comes, T., & Lauras, M. (2021). Designing a framework for designing humanitarian blockchain projects. *Computers in Industry, 131*(2021). https://doi.org/10.1016/j.compind.2021.103487

Bailey, D. E., & Leonardi, P. M. (2015). *Technology choices: Why occupations differ in their embrace of new technology*. MIT Press.

Balci, O. (1994). Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Annals of Operations Research*, *53*, 121–173. https://doi.org/10.1007/BF02136828.

Baldwin, C.Y. & Clark, K.B. (2000). *Design Rules: The power of modularity*. MIT Press.

Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press.

BCG & NASSCOM. (2022). Sandboxing into the future: Decoding technology's biggest bets. Boston Consulting Group and National Association of Software and Service Companies. Retrieved from https://www.bcg.com/publications/2022/india-sandboxing-into-the-future-decoding-technologys-biggest-bets [06 October 2023].

Benbya, H. & McKelvey, B. (2006). Toward complexity theory of information system development. *Information, Technology and People, 19*(1), 12–34. https://doi.org/10.1108/09593840610649952.

Benbya, H., Nan, N., Tanriverdi, H., & Yoo, Y. (2020). Complexity and information systems research in the emerging digital world. *MIS Quarterly, 44*(1), 1-17. https://doi.org/10.25300/MISQ/2020/13304.

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. Yale University Press.

Bessant, J., Ramalingam, B., Rush, H., Marshall, N., Hoffman, K., & Gray, B. (2014). *Innovation management, innovation ecosystems and humanitarian innovation*. Literature Review for the Humanitarian Innovation Ecosystem Research Project. Retrieved from https://bit.ly/3M7Ibh0 [29 March 2023].

Boland, R.J., Lyytinen, K. & Yoo, Y. (2007). Wakes of innovation in project networks: The case of digital 3-D representations in architecture, engineering and construction. *Organization Science, 18*(4), 631–647. https://doi.org/10.1287/orsc.1070.0304.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. P*roceedings of National Academy of Sciences, 99*(3), 7280–7287. https://doi.org/10.1073/pnas.082080899.

Bonina, C., Koskinen, K., Eaton, B., & Gawer, A. (2021). Digital platforms for development: Foundations and research agenda. *Information Systems Journal, 31*(6), 869-902. https://doi.org/10.1111/isj.12326.

Borshchev, A. & Filippov, A. (2004). From system dynamics to agent-based modeling. *The 22nd International Conference of the System Dynamics Society*, 25–29.

Boudreau, K. J. (2012). Let a thousand flowers bloom? An early look at large numbers of software app developers and patterns of innovation. *Organization Science*, *23*(5), 1409–1427. http://www.jstor.org/stable/23252315.

Buolamwini, J. & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research, 81, 77-91. https://proceedings.mlr.press/v81/buolamwini18a.html.

Burns, R. (2018). Datafying disaster: Institutional framings of data production following superstorm sandy. *Annals of the American Association of Geographers, 108*(2), 569-578. https://doi.org/10.1080/24694452.2017.1402673.

Brynjolfsson, E. & Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic Perspectives*, *14*(4), 23-48. https://doi.org/10.1257/jep.14.4.23.

Campagnolo, D. & Camuffo, A. (2009). The concept of modularity in management studies: A literature review. *International Journal of Management Reviews*, *12*(3), 259–283. https://doi.org/10.1111/j.1468-2370.2009.00260.x.

Cartwright, N. (2004). Causation: One word, many things. *Philosophy of Science*, *71*(5), 805-819. https://doi.org/10.1086/426771.

Castle, C. & Crooks, A. (2006). *Principles and concepts of agent-based modelling for developing geospatial simulations*. Technical report. Centre for Advanced Spatial Analysis, University College London.

Cecez-Kecmanovic, D., Galliers, R. D., Henfridsson, O., Newell, S., & Vidgen, R. (2014). The sociomateriality of information systems: Current status, future directions. *MIS Quarterly, 38*(3), 809-830. https://doi.org/10.25300/MISQ/2014/38:3.3.

Chorafas, D.N. (1965). *Systems and simulation.* Academic Press Inc.

Chui, M., Issler, M., Roberts, R., & Yee, L. (2023). *McKinsey technology trends outlook: 2023*. McKinsey. Retrieved from https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-top-trends-in-tech#/ [28 August 2023].

Cirillo, P. & Gallegati, M. (2012). The empirical validation of an agent-based model. *Eastern Economic Journal, 38*(4), 525–547. https://doi.org/10.1057/eej.2011.34.

Clark, K.B. (1985). The interaction of design hierarchies and market concepts in technological evolution. *Research Policy, 14*(5), 235–251. https://doi.org/10.1016/0048-7333(85)90007-1.

Codastefano, M. (2019). *Machine learning in emergency response*. The World Food Programme. Retrieved from https://www.itu.int/dms_pub/itu-d/oth/07/1d/D071D0000050001PDFE.pdf [07 April 2024].

Collins, A. J. & Frydenlund, E. (2016). Agent-based modelling and strategic group formation: A refugee case study. *Proceedings of the 2016 Winter Simulation Conference*, 1289–1300. https://doi.org/10.1109/WSC.2016.7822184.

Constantinides, P. & Barrett, M. (2006). Large scale ICT innovation, power, organizational change: The case of a regional health information network. *The Journal of Applied Behavioral Science*, *42*, 76–90. https://doi.org/10.1177/002188630528429.

Constantinides, P., Henfridsson, O., & Parker, G.C. (2018). Introduction: Platforms and infrastructures in the digital age. *Information Systems Research*, *29*(2), 381-400. https://doi.org/10.1287/isre.2018.0794.

Costanza-Chock, S. (2020). *Design Justice*. MIT Press.

Crooks, A., Castle, C., & Batty, M. (2008). Key challenges in agent-based modelling for geo-spatial simulation. *Computers, Environment and Urban Systems*, *32*(6), 417–430. https://doi.org/10.1016/j.compenvurbsys.2008.09.004.

Currie, W.L. & Guah, M.W. (2007). Conflicting institutional logics: a national programme for IT in the organisational field of healthcare. *Journal of Information Technology*, *22*, 235-247. https://doi.org/10.1057/palgrave.jit.2000102.

Danermark, B., Ekström, M., Jakobsen, L., & Karlsson, J. (2002). *Explaining society: Critical realism in the social sciences*. Routledge.

Darking, M., Whitley, E.A., & Dini, P. (2008). Governing diversity in the digital ecosystem. *Communications of the ACM*, *51*(8), 137–140. https://doi.org/10.1145/1400181.1400211.

Davidsson, P., Holmgren, J., Kyhlbäck, H., Mengistu, D. and Persson, M. (2007). Applications of agent based simulation. In L. Antunes & K. Takadama (Eds). *Multi-Agent Based Simulation VII. MABS 2006*, Vol. 4442 Lecture Notes in Computer Science (15-27). Springer.

Davies, A. & Hobday, M. (2005). *The business of projects: Managing innovation in complex products and systems*. Cambridge University Press.

Davis, G. F. & Marquis, C. (2005). Prospects for organization theory in the early twenty-first century: Institutional fields and mechanisms. *Organization Science*, *16*(4), 332-343. https://doi.org/10.1287/orsc.1050.0137.

Davison, R. M. & Martinsons, M. G. (2016). Context is king! Considering particularism in research design and reporting. *Journal of Information Technology*, *31*(3), 241-249. https://doi.org/10.1057/jit.2015.19.

Dawes, S. S., Vidiasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, *33*(1), 15–27. https://doi.org/10.1016/j.giq.2016.01.003.

Dedehayir, O., Makinen, S.J., & Roland Ortt, J. (2018). Roles during innovation ecosystem genesis: a literature review. *Technological Forecasting and Social Change*, *136*, 18–29. https://doi.org/10.1016/J.TECHFORE.2016.11.028.

deReuver, M., Sorensen, C., & Basole, R. C. (2018). The digital platform: A research agenda. *Journal of Information Technology*, *33*(2), 124–135. https://doi.org/10.1057/s41265-016-0033-3.

Devaraj, S., & Kohli, R. (2003). Performance impacts of information technology: Is actual usage the missing link? *Management Science*, *49*(3), 273-289. https://www.jstor.org/stable/4133926.

Djeffal, C. (2023). *The EU AI Act at a crossroads: Generative AI as a challenge for regulation*. European Law Blog. Retrieved from https://europeanlawblog.eu/2023/07/24/the-eu-ai-act-at-a-crossroads-generative-ai-as-a-challenge-for-regulation/ [31 October 2023].

Doherty, N.F. & King, M. (2005). From technical to socio-technical change: tackling the human and organizational aspects of system development projects. *European Journal of Information Systems*, *14*(1), 1–5. https://doi.org/10.1057/palgrave.ejis.3000517.

Dooley, K. (1996). A nominal definition of complex adaptive systems. *The Chaos Network*, *8*(1), 2-3.

DRC (Danish Refugee Council). (n.d.). *Foresight: Displacement forecasts*. Retrieved from https://pro.drc.ngo/what-we-do/innovation-and-climate-action/predictive-analysis/foresight-displacement-forecasts/ [07 April 2024].

Duffield, M. (2015). The digital development-security nexus: Linking cyber humanitarianism and drone warfare. In P. Jackson, *Handbook of international security and development* (pp. 80–94). Cheltenham: Edward Elgar.

Eaton, B., Elaluf-Calderwood, S., Sørensen, C., & Yoo, Y. (2015). Distributed tuning of boundary resources: The case of Apple's iOS service system. *MIS Quarterly*, *39*(1), 217-243. https://doi.org/10.25300/MISQ/2015/39.1.10.

Eisenhardt, K. M. (1989). Building theories from case study research. *The Academy of Management Review*, *14*(4), 532–550. https://doi.org/10.5465/amr.1989.4308385.

Eisenhardt, K.M. & Graebner, M.E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, *50*(1), 25–32. https://doi.org/10.5465/amj.2007.24160888.

Eisenmann, T., Parker, G., & van Alstyne, M. (2006). Strategies for two-sided markets. *Harvard Business Review*, *84*(10), 1-10.

Elsenbroich, C. (2012). Explanation in agent-based modelling: Functions, causality or mechanisms? *Journal of Artificial Societies and Social Simulation, 15*(3). https://doi.org/10.18564/jasss.1958.

Elwick, J. (2012). Layered history: Styles of reasoning as stratified conditions of possibility. *Studies in History and Philosophy of Science, 43*, 619-627. https://doi.org/10.1016/j.shpsa.2012.07.004.

Endriss, U. (2011). Computational social choice: Prospects and challenges. *Procedia Computer Science*, *7*, 68-72. https://doi.org/10.1016/j.procs.2011.12.022.

Engestrom, Y. (2004) Managing as argumentative history-making. In R. Boland & F. Collopy (Eds). *Managing as designing* (96–101). Stanford University Press.

Entwisle, B., Williams, N. E., Verdery, A. M., Rindfuss, R. R., Walsh, S. J., Malanson, G. P., Mucha, P. J., Frizzelle, B. G., McDaniel, P. M., Yao, X., Heumann, B. W., Prasartkul, P., Sawangdee, Y. & Jampaklay, A. (2016). Climate shocks and migration: An agent-based modeling approach. *Population and Environment*, *38*(1), 47–71. https://doi.org/10.1007/s11111-016-0254-y.

Epstein, J. M. (2008). Why model? *Journal of Artificial Societies and Social Simulation, 11*(4). https://www.jasss.org/11/4/12.html.

Epstein, J. M. & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom up*. MIT Press.

Eskelund, K., Heeks, R., & Nicholson, B. (2019). Exploring an impact sourcing Initiative for a community of people with disabilities: A capability analysis. In P. Nielsen, P., H.C. Kimaro (Eds.). *Information and communication technologies for development. Strengthening southern-driven cooperation as a catalyst for ICT4D*. ICT4D 2019. IFIP Advances in Information and Communication Technology, Vol 551. Springer. https://doi.org/10.1007/978-3-030-18400-1_5.

Evans, D.S., Hagiu, A., & Schmalensee, R. (2006). *Invisible Engines: How software platforms drive innovation and transform industries*. MIT Press.

Faik, I., Barrett, M., & Oborn, E. (2020). How information technology matters in societal change: An affordance-based institutional logics perspective. *MIS Quarterly*, *44*(3), 1359-1390. https://doi.org/10.25300/MISQ/2020/14193.

Falleti, T. G. & Lynch, J. F. (2009). Context and causal mechanisms in political analysis. *Comparative political studies, 42*(9), 1143-1166. https://doi.org/10.1177/0010414009331724.

Faraj, S., & Azad, B. (2012). The materiality of technology: An affordance perspective. In P. M. Leonardi, B. A. Nardi, & J. Kallinikos (Eds). *Materiality and Organizing* (237-259). Oxford University Press.

Faraj, S., Jarvenpaa, S.L., & Majchrzak, A. (2011). Knowledge collaboration in online communities. *Organisation Science, 22*(5),1224–1239. https://doi.org/10.1287/orsc.1100.0614.

Faraj, S., von Krogh, G., Monteiro, E., & Lakhani, K.R. (2016). Special section introduction: Online community as space for knowledge flows. *Information Systems Research, 27*(4), 668–684. https://doi.org/10.1287/isre.2016.0682.

Faulkner, P. & Runde, J. (2009). On the identity of technological objects and user innovations in function. *Academy of Management Review*, *34*(3), 442-462. https://doi.org/10.5465/AMR.2009.40632318.

Faulkner, P. & Runde, J. (2013). Technological objects, social positions, and the transformational model of social activity. *MISQ, 37*(3), 803–818. https://www.jstor.org/stable/43826001.

Faulkner, P. & Runde, J. (2019). Theorizing the digital object. *MISQ*, *43*(4), 1-24. https://doi.org/10.17863/CAM.37903.

Fayard, A. L. & Weeks, J. (2007). Photocopiers and water-coolers: The affordances of informal interaction. *Organization Studies, 28*(5), 605-34. https://doi.org/10.1177/0170840606068310.

Feyeraband, P. (1975). *Against method: Outline of an anarchist theory of knowledge.* New Left Books.

Fichman, R.G., Dos Santos, B.L., & Zheng, Z. (2014). Digital innovation as a fundamental and powerful concept in the information systems curriculum. *MIS Quarterly, 38*(2), 329-354. https://doi.org/10.25300/MISQ/2014/38.2.01.

Fjellström, D. & Guttormsen, D. S. A. (2016). A critical exploration of "access" in qualitative international business field research: Towards a concept of socio-cultural and multidimensional research practice. *Qualitative research in organisations and management*, *11*(2), 110-126. https://doi.org/10.1108/QROM-05-2014-1225.

Flyvbjerg, B. (2011). Case study. In N. K. Denzin & Y. S. Lincoln (Eds). *The Sage Handbook of Qualitative Research* (301-316). Sage.

Fürstenau, D., Baiyere, A., Schewina, K., Schulte-Althoff, M., & Rothe, H. (2023). Extended generativity theory on digital platforms. *Information Systems Research, 34*(4), 1686-1710. https://doi.org/10.1287/isre.2023.1209.

Gandy, O.H. (2021). *The panoptic sort: A political economy of personal information.* Oxford University Press.

Garfinkel, A. (1981). *Forms of explanation: Rethinking the questions in social theory*. Yale University Press.

Garud, R. & Kumaraswamy, A. (1995). Technological and organisational designs for realizing economies of substitution. *Strategic Management Journal*, *16*, 93–109. https://doi.org/10.1002/smj.4250160919.

Garud, R., Jain, S., & Tuertscher, P. (2008). Incomplete by design and designing for incompleteness. *Organization Studies*, *29*(3), 351–371. https://doi.org/10.1177/01708406070880.

Ghazawneh, A. & Henfridsson, O. (2013). Balancing platform control and external contribution in third-party development: The boundary resources model. *Information Systems Journal*, *23*(2), 173–192. https://doi.org/10.1111/j.1365-2575.2012.00406.x.

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Polity Press.

Golden-Biddle, K. & Locke, K. (1993). Appealing work: an investigation of how ethnographic texts convince. *Organization Science 4*(4), 595–616. https://doi.org/10.1287/orsc.4.4.595.

Grabowski, M., & Roberts, K. H. (1999). Risk mitigation in virtual organizations. *Organization Science*, *10*(6), 704–721. https://doi.org/10.1287/orsc.10.6.704.

Greenwood, R., Suddaby, R. & Hinings, C.R. (2002). Theorizing change: The role of professional associations in the transformation of institutionalised fields. *Academy of Management Journal 45*(1), 58–80. https://doi.org/10.2307/3069285.

Gross, N. (2009). A pragmatist theory on social mechanisms. *American Sociological Review, 74*(3), 358-379. https://doi.org/10.1177/000312240907400302.

Hacking, I. (2002). *Historical ontology*. Harvard University Press.

Hamid, U. Z. A. & Suoheimo, M. (2023). Product development challenges for emerging technologies and service design roles in addressing the issues. In U. Z. A. Hamid & M. Suoheimo (Eds). *Service design for emerging technologies product development*. Springer Series in Design and Innovation, Vol. 29. https://doi.org/10.1007/978-3-031-29306-1_2.

Hanseth, O. & Lyytinen, K. (2010). Design theory for dynamic complexity in information infrastructures: The case of building Internet. *Journal of Information Technology, 25*(1), 1–19. https://doi.org/10.1057/jit.2009.19.

Harrison, T. M., Pardo, T. A., & Cook, M. E. (2012). Creating open government ecosystems: A research and development agenda. *Future Internet, 4*(4), 900–928. https://doi.org/10.3390/fi4040900.

Hassani-Mahmooei, B. & Parris, B. W. (2012). Climate change and internal migration patterns in Bangladesh: An agent-based model. *Environment and Development Economics*, *17*(06), 763–780. https://doi.org/10.1017/S1355770X12000290.

Hattle, A., Yang, K. S. and Zeng, S. (2016). Modelling the Syrian refugee crisis with agents and systems. *The UMAP Journal, 37*(2), 195–213.

Heath, B., Hill, R. & Ciarallo, F. (2009). A survey of agent-based modeling practices (January 1998 to July 2008). *Journal of Artificial Societies and Social Simulation, 12*(4). https://www.jasss.org/12/4/9.html.

Hébert, G., Perez, L. & Harati, S. (2018). An agent-based model to identify migration pathways of refugees: The case of Syria. In L. Perez, R. Sengupta, & E. Kim (Eds). *Agent-based models and complexity science in the age of geospatial big data* (45-58). https://doi.org/10.1007/978-3-319-65993-0_4.

Hedström, P. (2005). *Dissecting the social: On the principles of analytical sociology*. Cambridge University Press.

Hedström, P. & Swedberg, R. (1998). Social mechanisms: An introductory essay. In P. Hedström & R. Swedberg (Eds). *Social mechanisms: An analytical approach to social theory* (1-31). Cambridge University Press.

Henfridsson, O., Mathiassen, L., & Svahn, F. (2014). Managing technological change in the digital age: The role of architectural frames. *Journal of Information Technology, 29*(1), 27–43. https://doi.org/10.1057/jit.2013.30.

Heracleous, L. & Barrett, M. (2001). Organizational change as discourse: Communicative actions and deep structures in the context of information technology implementation. *Academy of Management Journal*, *44*, 755-778. https://doi.org/10.2307/3069414.

Hilhorst, D. & Jansen, B. J. (2010). Humanitarian space as arena: A perspective on the everyday politics of aid. *Development and Change, 41*(6), 1117–1139. https://doi.org/10.1111/j.1467-7660.2010.01673.x.

Hobday, M. (1998). Product complexity, innovation and industrial organisation. *Research Policy, 26*(6), 689–710. https://doi.org/10.1016/S0048-7333(97)00044-9.

Hobday, M., Prencipe, A., & Davies, A. (2003). Introduction. In A. Prencipe, A. Davies, & M. Hobday (Eds.). *The business of system integration* (1-14). Oxford University Press.

Hobday, M., Rush, H., & Tidd, J. (2000). Innovation in complex product and system. *Research Policy, 29*(7-8). https://doi.org/10.1016/S0048-7333(00)00105-0.

Holland, J. (1995). *Hidden order: How adaptation builds complexity.* Addison-Wesley.

Hund, A., Wagner, H., Beimborn, D., & Weitzel, T. (2021). Digital innovation: Review and novel perspective. *Journal of Strategic Information Systems, 30*(4), 1-39. https://doi.org/10.1016/j.jsis.2021.101695.

Iazzolino, G. (2023). ‘Going Karura’: Colliding subjectivities and labour struggle in Nairobi's gig economy. *Environment and Planning A: Economy and Space, 55*(5), 1114-1130. https://doi.org/10.1177/0308518X211031916.

ITU (International Telecommunications Union). (n.d.). *Digital technologies to achieve UN SDGs.* Retrieved from https://www.itu.int/en/mediacentre/backgrounders/Pages/icts-to-achieve-the-united-nations-sustainable-development-goals.aspx [28 August 2023].

Jacobides, M.G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal, 39,* 2255 – 2276. https://doi.org/10.1002/smj.2904.

Jacobsen, K.L. & Fast, L. (2019). Rethinking access: How humanitarian technology governance blurs control and care. *Disaster, 43*(S2), S151 - S168. https://doi.org/10.1111/disa.12333.

Jelonek, A.C. (2023). *Use of predictive analytics in humanitarian operations* (Dissertation, Fordham University). ETD Collection for Fordham University. Retrieved from https://research.library.fordham.edu/dissertations/AAI30246999/ [07 April 2024].

Johnson, R. T., Lampe, T. A. & Seichter, S. (2009). Calibration of an agent-based simulation model depicting a refugee camp scenario. *Proceedings of the 2009 Winter Simulation Conference*, 1778–1786. https://doi.org/10.1109/WSC.2009.5429180.

Jones, M. (1999). Structuration theory. In W. L. Currie & R. D. Galliers (Eds). *Rethinking management information systems: An interdisciplinary perspective* (103–135). Oxford University Press.

Joseph D., Pizzini, G., & Pynnonen, L. (2020). *Images from above: A review of aerial assessment in support of humanitarian response operations.* American Red Cross and IFRC. Retrieved from https://americanredcross.github.io/images-from-above/index.html [07 April 2024].

Kallinikos, J. (2004). Deconstructing information packages: Organizational and behavioral implications of ERP systems. *Information Technology & People, 17*(1), 8-30. https://doi.org/10.1108/09593840410522152.

Kallinikos, J. (2006). *The consequences of information*. Edward Elgar Publishing.

Kallinikos, J. (2009). On the computational rendition of reality: Artefacts and human agency. *Organization, 16*(2), 183-202. https://doi.org/10.1177/1350508408100474.

Kallinikos, J. & Hasselbladh, H. (2009). Work, control and computation: Rethinking the legacy of neo-institutionalism. In R.E. Meyer, K. Sahlin, M.J. Ventresca, & Walgenbach, P. (Eds.). *Institutions and Ideology* (257-282). Emerald Group Publishing Limited.

Kallinikos, J., Aaltonen, A., & Marton, A. (2013). The ambivalent ontology of digital artifacts. *MIS Quarterly, 37*(2), 357-370. https://doi.org/10.25300/MISQ/2013/37.2.02.

Katz, M. & Shapiro, C. (1994). Systems competition and network effects. *Journal of Economic Perspectives*, *8*(2), 93-115. https://doi.org/0.1257/jep.8.2.93.

Klabunde, A., & Willekens, F. (2016). Decision-making in agent-based models of migration: State of the art and challenges. *European Journal of Population, 32*(1), 73–97. https://doi.org/10.1007/s10680-015-9362-0 .

Klein, H.K. & Myers, M.D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly*, *23*(1), 67–93. https://doi.org/10.2307/249410.

Kleine, D. (2007). *Empowerment and the limits of choice: Microentrepreneurs, information and communication technologies and state policies in Chile* (PhD thesis, London School of Economics). https://scholar.google.co.uk/citations?view_op=view_citation&hl=en&user=KAKPvgQAAAAJ&citation_for_view=KAKPvgQAAAAJ:4DMP91E08xMC

Kleine, D. (2010). ICT4WHAT?—Using the choice framework to operationalise the capability approach to development. *Journal of International Development*, *22*, 674-692. https://doi.org/10.1002/jid.1719.

Kleine, D. (2011). The capability approach and the 'medium of choice': steps towards conceptualising information and communication technologies for development. E*thics and Information Technology, 13,* 119–130. https://doi.org/10.1007/s10676-010-9251-5.

Kling, R. (1987). Defining the boundaries of computing across complex organizations. In R. J. Boland Jr., & R. A. Hirschheim (Eds). *Critical issues in information systems research* (307-362). John Wiley.

Kling, R. (1992). Behind the Terminal: The critical role of computing infrastructure in effective information systems development and use. In W. Cotterman & J. Senn (Eds.) *Challenges and Strategies for Research in Systems Development* (153-201). John Wiley.

Kling, R. & Scacchi, W. (1982). The web of computing: Computing technology and social organization. *Advances in Computers, 21*, 1-90. https://doi.org/10.1016/S0065-2458(08)60567-7.

Knights, D. & Murray, F. (1994). *Managers divided: Organisation politics and information technology management*. Wiley.

Kniveton, D. R., Smith, C. D. & Black, R. (2012). Emerging migration flows in a changing climate in dryland Africa. *Nature Climate Change*, *2*(6), 444–447. https://doi.org/10.1038/nclimate1447.

Kniveton, D., Smith, C. & Wood, S. (2011). Agent-based model simulations of future changes in migration flows for Burkina Faso. *Global Environmental Change, 21*, S34–S40. https://doi.org/10.1016/j.gloenvcha.2011.09.006.

Knorr-Cetina, K. (1981). The micro-sociological challenge of macro-sociology: Towards a reconstruction of social theory and methodology. In K. Knorr-Cetina & A. V. Cicourel (Eds). *Advances in social theory and methodology* (1-47). Routlege and Kegan Paul.

Koskela-Huotari, K., Edvardsson, B., Jonas, J.M., Sörhammar, D., & Witell, L. (2016). Innovation in service ecosystems—breaking, making, and maintaining institutionalized rules of resource integration. *Journal of Business Resources, 69*, 2964–2971. https://doi.org/10.1016/j.jbusres.2016.02.029.

Kumar, P. (2023). *Data evaluation for LLM: Enhancing accuracy & responsibility*. Future Bee AI Data Store. Retrieved from https://www.futurebeeai.com/blog/data-evaluation-for-llm-enhancing-accuracy-and-responsibility [31 October 2023].

Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. Basic Books.

Lamb, R. & Kling, R. (2003). Reconceptualizing users as social actors in information systems research. *MIS Quarterly, 27*(2), 197-235. https://doi.org/10.2307/30036529.

Langley, A. (1999). Strategies for theorising from process data. *The Academy of Management Review, 24*(4), 691-710. https://doi.org/10.2307/259349.

Langlois, R.N. & Robertson, P.L. (1992). Networks and innovation in a modular system: Lessons from the microcomputer and stereo component industries. *Research Policy, 21*(4), 297–313. https://doi.org/10.1016/0048-7333(92)90030-8.

Langlois, R.N. (2003). The vanishing hand: The changing dynamics of industrial capitalism. *Industrial and Corporate Change, 12*(2), 351-385. https://doi.org/10.1093/icc/12.2.351.

Latek, M. M., Mussavi Rizi, S. M. & Geller, A. (2013). Verification through calibration: An approach and a case study of a model of conflict in Syria. *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World*, 1649–1660. IEEE Press. https://doi.org/10.1109/WSC.2013.6721547.

Latorre-Catalán, M. (2017). Opportunity structure. In The Wiley-Blackwell Encyclopedia of Social Theory (1-2). Https://doi.org/10.1002/9781118430873.est0268.

Latour, B. (2005). *Reassembling the Social: An introduction to Actor–Network Theory*. Oxford University Press.

Law, A. M. (2008). How to build valid and credible simulation models. In S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson & J. W. Fowler (Eds). *Proceedings of the 2008 Winter Simulation Conference*, 39–47.

Law, J. (1991). *A Sociology of monsters: Essays on power, technology and domination*. Routledge.

Leavitt, H. J. (1964). Applied organization change in industry: structural, technical, and human approaches. In S. Cooper, H. J. Leavitt, K. Shelly (Eds). *New perspectives in organizational research* (55–71). Wiley.

Lee, A. S. & Baskerville, R. L. (2003). Generalizing generalizability in information systems research. *Information Systems Research, 14*(3), 221–243. https://doi.org/10.1287/isre.14.3.221.16560.

Leonard-Barton, D. & Kraus, W. A. (1985). Implementing new technology. Harvard Business Review. Retrieved from https://hbr.org/1985/11/implementing-new-technology [06 October 2023].

Leonardi, P. M. (2011). When flexible routines meet flexible technologies: Affordance, constraint, and the Imbrication of human and material agencies. *MIS Quarterly*, *35*(1), 147-167. https://doi.org/10.2307/23043493.

Leonardi, P. M. (2012). Materiality, sociomateriality, and socio-technical systems: What do these terms mean? How are they different? Do we need them. In P. M. Leonardi, B. A. Nardi, & J. Kallinikos (Eds.). *Materiality and organizing: Social interaction in a technological world* (25-48). Oxford, UK: Oxford University Press.

Leonardi, P.M., Woo, D., & Barley, W.C. (2021). On the making of crystal balls: Five lessons about simulation modelling and the organization of work. *Information and Organisation*, *31*(2021), 1-8. https://doi.org/10.1016/j.infoandorg.2021.100339.

Lessig, L. (2000). Code and other laws of cyberspace. New York: Basic Books.

Lin, L., Carley, K. M. & Cheng, S. (2016). An agent-based approach to human migration. MIT Press.

Lund, C. (2014). Of what is this a case?: Analytical movements in qualitative social science research. *Human Organization*, *73*(3), 224 – 234. https://doi.org/10.17730/humo.73.3.e35q482014x033l4.

Lyyra, A.K. (2018). *From components to compositions: (De-)construction of computer-controlled behaviour with the Robot Operating System* (PhD thesis, London School of Economics and Political Science). LSE Theses Online. https://etheses.lse.ac.uk/3837/.

Lyytinen, K. & Newman, M. (2008). Explaining information systems change: a punctuated socio-technical change model. *European Journal of Information Systems*, *17*, 589–613, https://doi.org/10.1057/ejis.2008.50.

Lyytinen, K.J., Yoo, Y., & Boland, R.J. (2016). Digital product innovation within four classes of innovation networks. *Information Systems Journal*, *26*, 47 - 75. https://doi.org/10.1111/isj.12093.

Macal, C. M. & North, M. J. (2010). Tutorial on agent-based modelling and simulation. *Journal of Simulation*, *4*(3), 151–162. https://doi.org/10.1057/jos.2010.3.

Madon, S. & Schoemaker, E. (2021). Digital identity as a platform for improving refugee management. *Information Systems Journal*, *31*(6). https://doi.org/10.1111/isj.12353.

Madon, S. (1992). Computer based information systems for development planning. In S. C. Bhatnagar & M. Odedra (Eds). *Social implications of computers in developing countries* (209-217). Tata McGraw-Hill.

Majchrzak, A., Markus, M. L., & Wareham, J. (2016). Designing for digital transformation: Lessons for information systems research from the study of ICT and societal challenges. *MIS Quarterly*, *40*(2), 267–277.

Markus, L. & Nan, W. (2020). Theorizing connections between digital innovation and societal transformation: Learning from the case of M-Pesa in Kenya. In S. Nambisan, K. Lyytinen, & Y. Yoo (Eds.). *Handbook of digital innovation* (64-82). Edward Elgar Publishing.

Markus, M. L. & Robey, D. (1988). Information technology and organizational change: Causal structure in theory and research. *Management Science, 34*(5), 583-598. https://doi.org/10.1287/mnsc.34.5.583.

Markus, M. L. & Rowe, F. (2018). Is IT changing the world? Conceptions of causality for information systems theorizing. *MIS Quarterly, 42*(4), 1255-1280. https://doi.org/10.25300/MISQ/2018/12903.

Markus, M. L. & Silver, M. S. (2008). A foundation for the study of IT effects: A new look at DeSanctis and Pooles's concepts of structural features and spirit. *Journal of the Association for Information systems, 9*(10). https://doi.org/10.17705/1jais.00176.

Martin, M. (2011). Geertz and the interpretive approach in Anthropology. In A. P. Vayda & B. B. Walters (Eds). *Causal explanation for social scientists* (56-74). Rawman & Littlefield.

Marton A., Avital, M., & Jensen T.B. (2013). Reframing open big data. In *21st European conference on information systems*, Utrecht 5–8 June. https://research.cbs.dk/en/publications/reframing-open-big-data.

McKinsey. (2023). *Unveiling the next frontier of engineering simulation*. Retrieved from https://www.mckinsey.com/capabilities/operations/our-insights/unveiling-the-next-frontier-of-engineering-simulation [22 August 2023].

McLeod, L. & Doolin, B. (2012). Information systems development as situated socio-technical change: a process approach. *European Journal of Information Systems, 21*(2), 176-191. https://doi.org/10.1057/ejis.2011.43.

Miller, R., Hobday, M., Leroux-Demers, T., & Olleros, X. (1995). Innovation in complex systems industries: The case of flight simulation. *Industrial and Corporate Change, 4*(2), 363–400. https://academic.oup.com/icc/article-abstract/4/2/363/749729.

Mills, A. J., Durepos, G., & Weibe, E. (2010). *Encyclopedia of case study research*. Sage.

MIT Technology Review Insights. (2023). *Successfully deploying machine learning.* In association with JP Morgan Chase. Retrieved from https://www.technologyreview.com/2023/06/12/1074272/successfully-deploying-machine-learning/ [24 August 2023].

Molnar, P. (2020). *Technological testing grounds: Migration management experiments and reflections from the ground up*. EDRi and The Refugee Law Lab. Retrieved from https://edri.org/wp-content/uploads/2020/11/Technological-Testing-Grounds.pdf [28 March 2023].

Molnar, P. & Gill, L. (2018). *Bots at the gate: A human rights analysis of automated decision-making in Canada's immigration and refugee system*. University of Toronto International Human Rights Program and the Citizen Lab. https://citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf.

Moshtari, M. (2016). Inter-organizational fit, relationship management capability, and collaborative performance within a humanitarian setting. *Production and Operations Management*, *25*(9), 1542-1557. https://doi.org/10.1111/poms.12568.

Mumford, E. (2000). A socio-technical approach to systems design. *Requirements Engineering, 5*, 125–133. https://doi.org/10.1007/PL00010345.

Murmann, J.P. & Frenken, K. (2006). Toward a systematic framework for research on dominant designs, technological innovations, and industrial change. *Research Policy*, *35*(7), 925–952. https://doi.org/10.1016/j.respol.2006.04.011.

Nachmias, C. & Nachmias, D. (1992). *Research methods in the social sciences*. St. Martin's Press.

Najibi, A. (2020). *Racial discrimination in face recognition technology*. Harvard University Science in the News. Retrieved from https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/ [06 October 2023].

Nambisan, S. (2018). Architecture vs. ecosystem perspectives: Reflections on digital innovation. *Information and Organization, 28*(2018), 104 – 106. https://doi.org/10.1016/j.infoandorg.2018.04.

Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital innovation management: Reinventing innovation management research in a digital world. *MIS Quarterly, 41*(1), 223–238. https://doi.org/10.25300/MISQ/2017/41:1.03.

Nielsen, B.F., Sandvik, K.B., & Jumbert, M.B. (2016). How can innovation deliver humanitarian outcomes? *PRIO Policy Brief, 12*. Peace Research Institute Oslo (PRIO).

Orlikowski, W. J. (1992). The duality of technology: rethinking the concept of technology in organizations. *Organization Science* 3(3), 398–427. https://doi.org/10.1287/orsc.3.3.398.

Orlikowski, W. J. & Iacono, S. C. (2001). Desperately Seeking the 'IT' in IT Research—A Call to Theorizing the IT Artifact. *Information Systems Research, 12*(2), 121-134.

Orlikowski, W. J. & Scott, S. V. (2008). Sociomateriality: Challenging the separation of technology, work and organization. *The Academy of Management Annals, 2*(1), 433-474. https://doi.org/10.5465/19416520802211644.

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying information technology in organizations: Research approaches and assumptions. *Information Systems Research, 2*(1), 1-28. https://www.jstor.org/stable/23010611.

Parker, G., van Alstyne, M.W., & Choudary, S.P. (2016). Platform revolution: How networked markets are transforming the economy and how to make them work for you. Norton & Company.

Parnas, D.L. (1972). On the criteria to be used in decomposing systems into modules. *Communications of the ACM, 15*(12), 1053–1058. https://doi.org/10.1145/361598.361623.

Pentland, B.T. (1999). Building process theory with narrative: From description to explanation. *Academy of Management Review, 24*(4), 711-724. https://doi.org/10.2307/259350.

Pettigrew A.M., Woodman, R.W., & Cameron, K.S. (2001). Studying organizational change and development: challenges for future research. *Academy of Management Journal, 44*(4), 697–713. https://doi.org/10.5465/3069411.

Pettigrew, A. M. (1985). Contextualist research and the study of organisational change processes. In E. Mumford, R. Hirschheim, G. Fitzgerald & A. T. Wood-Harper (Eds). *Research methods in information systems* (53-78). North-Holland.

Pettigrew, A.M. (1987). Context and action in the transformation of firm. *Journal of Management Studies, 24*(6), 649–670. https://doi.org/10.1111/j.1467-6486.1987.tb00467.x.

Pettigrew, A. M. (1988). Longitudinal field research on change: Theory and practice. *Paper presented at the National Science Foundation Conference on Longitudinal Research Methods in Organizations, Austin.*

Pettigrew, A. M. (1990). Longitudinal field research on change: theory and practice. *Organization Science, 1*(3), 267–292. https://doi.org/10.1287/orsc.1.3.267.

Pettigrew, A. M. (1992). The character and significance of strategy process research. *Strategic Management Journal, 13*(S2), 5-16. https://doi.org/10.1002/smj.4250130903.

Pettigrew, A.M. (1997). What is processual analysis? *Scandinavian Journal of Management, 13*(4), 337–348. https://doi.org/10.1016/S0956-5221(97)00020-1.

Pham, K. & Luengo-Oroz, M. (2022). *Predictive modelling of movement of refugees and internally displaced people: Towards a computational framework*. arXiv, https://doi.org/10.48550/arXiv.2201.08006.

Pollock, N. & Williams, R. (2009). Software and organizations: The biography of the enterprise-eide system or how SAP conquered the world. Routledge.

Poveda, S. (2016). *Conscientisation and human development: The case of digital inclusion programmes in Brazil* (PhD thesis, Royal Holloway). https://pure.royalholloway.ac.uk/en/publications/conscientisation-and-human-development- the-case-of-digital-inclus.

Project Jetson. (n.d.) *Welcome to Project Jetson*. Project Jetson Wiki, Jetson GitHub. Retrieved from https://github.com/unhcr/Jetson/wiki [07 April 2024].

Ragin, C. C. (1987). *The comparative method*. University of California Press.

Ragin, C. C. (1992). "Casing" and the process of social inquiry. In C. C. Ragin & H. S. Becker (Eds). *What is a case? Exploring the foundations of social inquiry* (217–226). Cambridge University Press.

Raji, I. D. & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products, 429-435. AIES '19: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. https://doi.org/10.1145/3306618.3314244.

Raymer, J. & Smith, P. W. F. (2010). Editorial: Modelling migration flows. *Journal of the Royal Statistical Society, 173*(4), 703–705. https://www.jstor.org/stable/40925395.

Read, R., Taithe, B., & Mac Ginty, R. (2016). Data hubris? Humanitarian information systems and the mirage of technology. *Third World Quarterly, 37*(8), 1314-1331. https://doi.org/10.1080/01436597.2015.1136208.

Robey, D. & Newman, M. (1996). Sequential patterns in information systems development: An application of social process model. *ACM Transactions on Information Systems, 14*(1), 30-63. https://doi.org/10.1145/214174.214178.

Robey, D., Welke, R. & Turk, D. (2001). Traditional, iterative, and component-based development: A social analysis of software development paradigms. *Information Technology and Management*, *2*, 53-70. https://doi.org/10.1023/A:1009982704160.

Rush, H., Marshall, N., Bessant, J., & Ramalingam, B. (2021). Applying an ecosystems approach to humanitarian innovation. *Technological Forecasting and Social Change, 165*(2021). https://doi.org/10.1016/j.techfore.2020.120529.

Sabamurthy, V. & Zmud, R. W. (2000). The organizing logic for an enterprise's IT activities in the digital era - A prognosis of practice and a call for research. *Information Systems Research*, *11*(2), 105-114. https://doi.org/10.1287/isre.11.2.105.11780.

Sabherwal, R. & Newman, M. (2003). Persistence and change in system development: A dialectical view. *Journal of Information Technology*, *18*(2), 69-92. https://doi.org/10.1080/0268396032000101144.

Salmon, W. (1998). *Causality and explanation*. Oxford University Press.

Salvador, F. (2007). Toward a product system modularity construct: Literature review and reconceptualization. *IEEE Transactions on Engineering Management*, *54*(2), 219–240. https://doi.org/10.1109/TEM.2007.893996.

Sanchez, R. & Mahoney, J.T. (1996). Modularity, flexibility, and knowledge management in product and organization design. *Strategic Management Journal*, *17*(S2), 63–76. https://doi.org/10.1002/smj.4250171107.

Sandvik, K.B., Jumbert, M.G., Karlsrud, J., & Kaufmann, M. (2014). Humanitarian technology: A critical research agenda. *International Review of the Red Cross, 96*(893), 219 - 242. https://doi.org/10.1017/S1816383114000344.

Sargent, R. G. (2011). Verification and validation of simulation models. In S. Jain, R. Creasey, J. Himmelspach, K. White & M. Fu (Eds). *Proceedings of the 2011 Winter Simulation Conference*, 183–198.

Satariano, A. & Mozur, P. (2023). The people onscreen are fake. The disinformation is real. The New York Times. Retrieved from https://www.nytimes.com/2023/02/07/technology/artificial-intelligence-training-deepfake.html [06 October 2023].

Sawyer, S., Allen, J.P., & Lee, H. (2003). Broadband and mobile opportunities: a socio-technical perspective. *Journal of Information Technology*, *18*(2), 121–136. https://doi.org/10.1080/0268396032000101171.

Scharfstein, B.A. (1989). *The dilemma of context*. New York University Press.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology, 1*(2), 143–186. https://doi.org/10.1080/0022250X.1971.9989794.

Schilling, M.A. (2000). Toward a general modular systems theory and its application to interfirm product modularity. *The Academy of Management Review, 25*(2), 312–334. https://doi.org/10.5465/AMR.2000.3312918.

Scott, W.R., Ruef, M., Mendel, P.J., & Caronna, C.A. (2000). *Institutional Change and Healthcare Organisations*. University of Chicago Press.

Seawright, J. & Gerring, J. (2008). Case selection techniques in case study research. *Political Research Quarterly, 6*1(2), 294-308. https://doi.org/10.1177/1065912907313077.

Seddon, P.B. & Scheepers, R. (2012). Towards the improved treatment of generalization of knowledge claims in IS research: drawing general conclusions from samples. *European Journal of Information Systems*, *21*(1), 6–21. https://doi.org/10.1057/ejis.2011.9.

Sen, A. (1977). Social choice theory: A re-examination. *Econometrica*, *45*(1), 53-89.

Sen, A. (1984). *Resources, Values and Development*. Harvard University Press.

Sen, A. (1999). *Development as Freedom*. Oxford University Press: Oxford.

Shaikh, M. & Vaast, E. (2016). Folding and unfolding: Balancing openness and transparency in open source communities. *Information Systems Research*, *27*(4), 813–833. https://doi.org/10.1287/isre.2016.0646.

Simon, H. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, *106*(6), 467–482. https://www.jstor.org/stable/985254.

Simon, H. A. (1992). What is an "explanation" of behavior? *Psychological Science*, *3*(3), 150-161. https://doi.org/10.1111/j.1467-9280.1992.tb00017.x.

Sokolowski, J. A. & Banks, C. M. (2014). A methodology for environment and agent development to model population displacement. *Proceedings of the 2014 Symposium on Agent Directed Simulation, Society for Computer Simulation International*.

Spradley, J. P. (1980). *Participant observation*. Holt, Rinehart and Winston.

Svahn, F & Henfridsson, O. (2012). The dual regimes of digital innovation management. *2012 45th Hawaii International Conference on System Sciences*, 3347-3356. https://doi.org/10.1109/HICSS.2012.560.

Teece, D. (1993). The dynamics of industrial capitalism: Perspectives on Alfred Chandler's scale and scope. *Journal of Economic Literature, 31*(1), 199-225. http://www.jstor.org/stable/2728154.

Teece, D.J. (2018) Profiting from innovation in the digital economy: Enabling technologies, standards, and licensing models in the wireless world. *Research Policy, 47*(8), 1367–1387. https://doi.org/10.1016/j.respol.2017.01.015.

Tilly, C. (1984). *Big structures, large processes, huge comparisons*. Russell Sage Foundation.

Tilson, D., Lyytinen, K., & Sorensen, C. (2010). Digital infrastructures: The missing IS research agenda. *Information Systems Research*, *21*(4), 748-759. https://doi.org/10.1287/isre.1100.0318.

Tiwana, A., Konsysnski, B., & Bush, A.A. (2010). Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. I*nformation Systems Research*, *21*(4), 675 - 687. https://doi.org/10.1287/isre.1100.0323.

Townley, B. (2008). *Reason's neglect: Rationality and organizing*. Oxford University Press.

Tucker, I. (2017). A white mask worked better': why algorithms are not colour blind. The Guardian. Retrieved from https://www.theguardian.com/technology/2017/may/28/joy-buolamwini-when-algorithms-are-racist-facial-recognition-bias [06 October 2023].

Ulrich, K.T. (1995). The role of product architecture in the manufacturing firm. *Research Policy*, *24*(3), 419–440. https://doi.org/10.1016/0048-7333(94)00775-3.

Ulrich, K.T. & Eppinger, S.D. (2012). *Product design and development*. McGraw-Hill.

UN (United Nations). (n.d.). *Topics: Technology*. Retrieved from https://sdgs.un.org/topics/technology [28 August 2023].

UNHCR. (n.d.). *Welcome to the SatCropper UNHCR service*. Jetson Technical Specifications. Retrieved from https://satcropper.unhcr.org/ [07 April 2024].

UNOCHA (United Nations Office for the Coordination for Humanitarian Affairs). (2021). *From digital promise to frontline practice: New an emerging technologies in humanitarian action.* Retrieved from https://reliefweb.int/report/world/digital-promise-frontline-practice-new-and-emerging-technologies-humanitarian-action [29 August 2023].

UNOCHA. (n.d). *Anticipatory action*. Retrieved from https://www.unocha.org/anticipatory-action [07 April 2024].

UNOCHA. (2022). *Philippines anticipatory action framework*. Retrieved from https://www.unocha.org/publications/report/philippines/philippines-anticipatory-action-framework-2022-revision [07 April 2024].

UNOCHA. (2023). *Anticipatory action framework: Bangladesh monsoon floods – 2023 version*. https://reliefweb.int/report/bangladesh/anticipatory-action-framework-bangladesh-monsoon-floods-2023-version [07 July 2024]

Utterback, J.M. & Abernathy, W.J. (1975). A dynamic model of process and product innovation. *Omega*, *3*(6), 639–656. https://doi.org/10.1016/0305-0483(75)90068-7.

Van de Ven, A. H. (2005). Running in packs to develop knowledge intensive technologies. *MIS Quarterly, 29*(2), 365-378. https://doi.org/10.2307/25148683.

Van de Ven, A. H. (1992). Suggestions for studying strategy process: a research note. *Strategic Management Journal, 13*(S1), 169-188. https://doi.org/10.1002/smj.4250131013.

Van de Ven, A. H. & Poole, M. S. (1995). Explaining development and change in organizations. *Academy of Management Review, 20*(3), 510-540. https://doi.org/10.5465/amr.1995.9508080329.

Van Kuiken, S. (2022). Tech at the edge: Trends reshaping the future of IT and business. McKinsey. Retrieved from https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-at-the-edge-trends-reshaping-the-future-of-it-and-business [05 October 2023].

Van Maanen, J. (1983). The fact of fiction in organizational ethnography. *Administrative Science Quarterly, 24*(4), 539-550. https://doi.org/10.2307/2392360.

Wajcman, J. (2004). Technofeminism. Oxford: Polity Press.

Walsham, G. (1993). *Interpreting information systems in organizations*. John Wiley.

Walsham, G. (1995). The emergence of interpretivism in IS research. *Information Systems Research*, *6*(4), 376-394. https://doi.org/10.1287/isre.6.4.376

Walsham, G. (2006). Doing interpretive research. *European Journal of Information Systems, 15*(3), 320-330. https://doi.org/10.1057/palgrave.ejis.3000589.

Wareham, J., Fox, P.B., & Giner, J.L.C. (2014). Technology ecosystem governance. *Organization Science*, *25*(4), 1195-1215. https://doi.org/10.1287/orsc.2014.0895.

Whelan, S. & Verity, A. (2023). *How earth observation can enable anticipatory action in humanitarian crises.* Digital Humanitarian Network. Retrieved from https://www.preventionweb.net/publication/how-earth-observation-can-enable-anticipatory-action-humanitarian-crises [07 April 2024].

Wiggers, K. (2023). *Researchers discover a way to make ChatGPT consistently toxic. TechCrunch*. Retrieved from https://tinyurl.com/378pc6cm [31 October 2023].

Williams, R. & Pollock, N. (2008). *Software and organisations – The biography of the enterprise-wide system or how SAP conquered the world*. Routledge.

Winter, S., Berente, N., Howison, J., & Butler, B. (2014). Beyond the organizational ‘container’: Conceptualizing 21st century sociotechnical work. *Information and Organization, 24*(2014), 250-269. https://doi.org/10.1016/j.infoandorg.2014.10.003.

Yin, R. K. (2009). *Case study research: design and methods*. Sage.

Yin, R. K. (2014). *Case study research: design and methods*. Sage.

Yin, R. K. & Davis, D. (2007). Adding new dimensions to case study evaluations: The case of evaluating comprehensive reforms. *New Directions for Evaluation, 2007*(113), 75-93. https://doi.org/10.1002/ev.216.

Yoo, Y. (2013). The tables have turned: How can the information systems field contribute to technology and innovation management research? *Journal of the Association for Information Systems*, *14*(5), 227-236. https://doi.org/10.17705/1jais.00334.

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). The new organizing logic of digital innovation: An agenda for information systems research. *Information Systems Research, 21*(4), 724-735. https://doi.org/10.1287/isre.1100.0322.

Zammuto, R. F., Griffith, T. L., Majchrzak, A., Dougherty, D. J., & Faraj, S. (2007). Information technology and the changing fabric of organization. *Organization Science*, *18*(5), 749-762. https://doi.org/10.1287/orsc.1070.0307.

Zelezny-Green, R. (2017). *The role of girls' mobile phone use to increase access to educational content after school: A capabilities-based evaluation in Nairobi* (PhD Thesis, Royal Holloway). https://pure.royalholloway.ac.uk/en/publications/the-role-of-girls-mobile-phone-use- to-increase-access-to-educatio.

Zittrain, J. (2008). *The future of the internet--And how to stop it*. Yale University Press.

Zittrain, J.L. (2006). The generative internet. *Harvard Law Review, 119*(7),1974–2040. http://www.jstor.org/stable/4093608.

Zucker, L.G. (1983). Organisations as Institutions. In S.B. Bacharach (Ed.). *Research in the Sociology of Organisations* (1-42). JAI Press.

9. APPENDICES

9.1 Appendix 1: Process timeline summarising key phases

2016	
Phase / Outcome (P1)	Initial model
Source	Groen, 2016
Conflict	Mali
Resources portfolio:	
Computational	Pandas (pandas.pydata.org) for data analysis, matpolib for visualisation
Data	UNHCR, Bing Maps
Knowledge	ABM framework, linear interpolation, validation through sum of absolute differences
Information	Secondary sources
Financial	-
Human	Single developer
Interpersonal relationships	Single developer
Structural conditions	2015-16 refugee crisis; Recounting of refugee numbers by UNHCR makes it difficult to validate
Resource created	Computational (Flee code)
Functionality	Predicting refugee arrival due to forced displacement
Operational choice	Proof of concept
2017	
Phase / Outcome (P2)	Generalisation
Source	Suleimenova et al., 2017a, 1st Review feedback
Conflicts	Burundi, CAR, Mali
Resource portfolio:	
Computational	Flee, pandas, numpy
Data	UNHCR, ACLED, Bing Maps
Knowledge	Simplified simulation development process (Heath et al., 2019), Average relative difference for validation, Mean Absolute Scaled Error (MASE)
Information	Intuition
Financial	-
Human	Team (DS joining as PhD student), Academic
Interpersonal relationships	Team, Academic (BUL)
Structural conditions	Need for rapid simulation construction in humanitarian context; Removed links when border closures reported by UNHCR
Resource created	Knowledge (SDA)
Functionality	Develop, run, and validate results for refugee movements
Operational choice	Operational relevance
Phase / Outcome (P3)	Automation
Source	Suleimenova et al. 2017b
Conflict	-
Resource portfolio:	
Computational	FabSim, CARTO map construction platform
Data	UNHCR, ACLED, Population databases (like CityPop), Bing Maps
Knowledge	SDA
Information	-
Financial	Conference and project funding (Science Hackathon Geneva 2017 (CERN), Collaborations Workshop 2017 (Software Sustainability Institute), ComPat project)
Human	Team, Academic
Interpersonal relationships	Academic, Colleagues from SSI (technical input)
Structural conditions	UNHCR portal and API documentations for older platform with some conflict still under older UNHCR API
Resource created	Computational (FabFlee)
Functionality	Automated simulation workflow (only automated construction of network maps at this stage)
Operational choice	Efficiency (reduction in response time, avoiding errors and inaccuracies due to manual construction) and reusability
Phase / Outcome (P4)	Gamification
Source	Estrada et al., 2017, 1st review feedback
Conflict	Burundi
Resource portfolio:	
Computational	Kivy
Data	UNHCR, geospatial
Knowledge	Robust Facility Location Problem (RFLP)
Information	-
Financial	-

Human	Team, Academic
Interpersonal relationships	Academic
Structural conditions	Lack of reciprocal interest
Resource created	-
Functionality	Video game interface for determining camp locations for aid deployment
Operational choice	Usability (lay person non-computational background) and efficiency (Reducing response time in resource allocation in live or unfolding crisis)
2018	
Phase / Outcome (P5)	Parallelisation and multiscale coupling
Source	Groen, 2018
Conflict	-
Resource portfolio:	
Computational	MPI4Py
Data	-
Knowledge	-
Information	-
Financial	ComPat project
Human	Single developer
Interpersonal relationships	UCL (ComPat)
Structural conditions	Incorporating additional causal conditions through small scale models, enhancing code performance
Resource created	Knowledge (multiscale simulation approach), Computational (Parallelised Flee; Coupling file)
Functionality	Parallel execution for better performance and coupling to additional models
Operational choice	Simplicity (over scalability)
Phase / Outcome (P6)	Input data feasibility
Source	Chan et al., 2018
Conflict	CAR
Resource portfolio:	
Computational	Tableau, Excel
Data	UNHCR, ACLED
Knowledge	SDA, Correlation between dependent (refugee count) and independent variables (type of conflict event, magnitude of event, fatalities)
Information	-
Financial	ComPat project
Human	Team, Academic
Interpersonal relationships	Academic (BUL)
Structural conditions	Infrequent update of refugee counts on the UNHCR portal, sources of data outside UNHCR are qualitative and in document format (i.e. not machine readable), multicausality behind movement of displaced people beyond input parameters above (border, camp, refugee registrations opening, families with children or elderly – lack of data), lack of granular data
Resource created	-
Functionality	Predict number of refugees based on events driving refugee movement
Operational choice	Robustness of approach (Exploring the strength of relationship between variable to develop multiscale simulation)
2019	
Phase / Outcome (P7)	Coupling – Food security sub-model
Source	Campos et al., 2019
Conflict	South Sudan
Resource portfolio:	
Computational	FabFlee
Data	UNHCR, ACLED, Bing Maps, IPC
Knowledge	SDA, spatial and temporal correlations
Information	Secondary sources
Financial	VECMA and HiDALGO
Human	Team, Academic
Interpersonal relationships	Team, Academic, VECMA, HiDALGO
Structural conditions	Including additional causal factors like food security for improving accuracy
Resource created	Computational ((New commands in FabFlee (running simulation including food security sub-model and generate graphs between two simulations))[4]
Functionality	Simulating the effects of food security on forced displacement
Operational choice	Accuracy
Phase / Outcome (P8)	Coupling – conflict evolution sub-model
Source	Groen et al., 2019a
Conflict	Mali
Resource portfolio:	
Computational	Eagle supercomputer (PSNC) (HiDALGO), FabFlee
Data	UNHCR, Bing Maps, ACLED, MSCI World Metals and Mining Index
Knowledge	Hybrid simulation modelling, SDA

---

[4] Food security sub-model is not listed as a resource created because it was not reused in subsequent development processes.

Information	Medicins sans Frontiers (on terminology), secondary sources
Financial	HiDALGO and VECMA
Human	Team (HA joined), Academic
Interpersonal relationships	Academic (BUL), HiDALGO, VECMA
Structural conditions	Multicausal nature of forced displacement and compounded causal effects
Resource created	-[5]
Functionality	Simulate the effect of conflict evolution on number of forcibly displaced people
Operational choice	Prototype
2020	
Phase / Outcome (P9)	Automated Ensemble Simulations
Source	Suleimenova & Groen, 2020
Conflict	South Sudan
Resource portfolio:	
Computational	FabFlee
Data	UNHCR, ACLED, Geospatial
Knowledge	SDA
Information	Secondary
Financial	HiDALGO and VECMA
Human	Team
Interpersonal relationships	Team, HiDALGO, VECMA
Structural conditions	Effect of policy decisions on refugee movements; Difficult to simulate due to lack of roads and food insecurity, lingering effect of policy decisions on journey times
Resource created	Knowledge (Automated policy exploration toolkit with integrated sensitivity analysis, extended SDA for counterfactual scenarios)
Functionality	Ensemble simulations and sensitivity analysis
Operational choice	Operational relevance (informing policy decisions)
2021	
Phase / Outcome (P10)	Multiscale migration modelling
Source	Jahani et al., 2021
Conflict	South Sudan
Resource portfolio:	
Computational	MUSCLE 3 , Eagle supercomputer
Data	ECMWF Climate Data Store, GloFAS Data (Copernicus project), UNHCR, ACLED, Geospatial
Knowledge	SDA, Multiscale simulation approach, File I/O, Acyclic (one-way coupling), hybrid simulation approach
Information	Secondary
Financial	HiDALGO
Human	Team (AJ joined), HiDALGO, Expertise on MUSCLE3 (Lourens Veen from the Dutch e-Science Centre)
Interpersonal relationships	Team, HiDALGO
Structural conditions	Effect of weather and route accessibility on distribution of forcibly displaced, Challenges in collection of input and validation data due incorporating additional causal relationships
Resource created	-
Functionality	Testing the effects of multiscale coupling (i.e. adding details) on Flee output
Operational choice	Accuracy, evaluating coupling approaches
Phase / Outcome (P11)	Scaling parallelised version of Flee
Source	Anastasiadis et al., 2021
Conflict	South Sudan
Resource portfolio:	
Computational	Hawk, Eagle, Vulcan supercomputers, pflee (python dependencies, numpy; scipy, MPI4Py)
Data	UNHCR, ACLED, Geospatial
Knowledge	Basic agent parallelisation, agent space parallelisation
Information	-
Financial	HiDALGO
Human	Team, HiDALGO
Interpersonal relationships	Team, HiDALGO
Structural conditions	Evaluate level of detail and visualisation on account of multicausality with implications for computational costs and efficiency
Resource created	-
Functionality	Scalability
Operational choice	Performance evaluation
Phase / Outcome (P12)	Automated location graph construction
Source	Schweimer et al., 2021, HiDALGO, 2021c
Conflict	CAR, South Sudan
Resource portfolio:	

---

[5] Flare sub-model is not listed as a resource created because it was not reused in subsequent development processes.

Computational	Rule based AI, Open Source Routing Machine from Open Street Maps (OSM), HPE Apollo SC (not within HiDALGO consortium)
Data	Open street maps
Knowledge	Triangle inequality, contraction hierarchies algorithm
Information	-
Financial	HiDALGO, Austrian COMET Program (KNOW Centre)
Human	Team, HiDALGO
Interpersonal relationships	HiDALGO
Structural conditions	Accuracy depends on existence of road network, road infrastructure, type of location (cities vs. villages)
Resource created	Computational (Route pruning algorithm)
Functionality	Automated construction of location graphs at scale
Operational choice	Speeding up visualisation (Manual location graph construction – time-consuming and error prone)
Phase / Outcome (P13)	Formalisation of Flee 2.0 rule set
Source	Suleimenova et al., 2021
Conflict	Burundi, South Sudan, CAR, Mali
Resource portfolio:	
Computational	VECMAtk (EasyVVUQ, QCG-PilotJob), FabFlee, Eagle supercomputer
Data	UNHCR, ACLED, Geospatial
Knowledge	Sensitivity analysis (Stochastic collocation and Sobol sensitivity indices), SDA
Information	IOM and NGOs on the field
Financial	VECMA and HiDALGO
Human	Team and VECMA
Interpersonal relationships	Team, VECMA, HiDALGO, IOM
Structural conditions	Understanding the simuland (the problem being simulated)
Resource created	Knowledge (SDSD), Computational (Flee rule set 2.0)
Functionality	Identifying pivotal parameters and rule set refinement
Operational choice	Accuracy (Ruleset refinement through identification of pivotal parameters)
2022	
Phase / Outcome (P14)	Multiobjective Optimisation (ML + ABM)
Source	Xue et al., 2022, Field Notes
Conflict	South Sudan
Resource portfolio:	
Computational	SDA, Evolutionary algorithm (NSGA-II), FabFlee, Route pruning algorithm
Data	UNHCR, ACLED, Geospatial
Knowledge	Facility Location Problem (FLP)
Information	-
Financial	ITFLOWS, HiDALGO
Human	Team, Academic, HiDALGO
Interpersonal relationships	Academic, HiDALGO, ITFLOWS (terminology – asylum seekers / unrecognised refugees)
Structural conditions	Incorporating construction and transportation costs
Resource created	Knowledge (Multiobjective simulation optimisation approach)
Functionality	Predicting optimal camp location optimising three objectives (min.travel distance, max.no.of people in camps, min.idle camp capacity)
Operational choice	Operational relevance (Finding optimal locations for camp placement for aid deployment)
Phase / Outcome (P15)	Integrated refugee IDP forecasting
Source	Suleimenova et al., 2022, Internal documents, Field Notes
Conflict	Tigray
Resource portfolio:	
Computational	Flee
Data	UNHCR, ACLED, OpenStreet Maps
Knowledge	SDA
Information	Qualitative reports, News (Associated Press, Reuters, BBC); Feedback from Save the Children
Financial	HiDALGO
Human	Team, Save the Children
Interpersonal relationships	Save the Children, HiDALGO
Structural conditions	Escalation of Tigray conflict; Data on displaced people not available for all locations
Resource created	-
Functionality	Predicting movement of refugees and IDPs
Operational choice	IDP simulation and forecasting

9.2 Appendix 2: Mapping the resource portfolio

	Computational	Informational	Data	Knowledge	Financial	Human
2016						
Initial model (P1)						
Resources required	Pandas (pandas.pydata.org) for data analysis, matpolib for visualisation	Secondary sources	UNHCR, Bing	ABM framework, linear interpolation, mathematical validation through sum of absolute differences	-	Single developer
Resources created	Flee	-	-	-	-	-
2017						
SDA (P2)						
Resources required	Flee, pandas, numpy	Intuition	UNHCR, ACLED, Bing Maps	Simplified simulation development process; Average relative difference; Mean Absolute Scale Error (MASE)	-	DS, Academic
Resources created	-	-	-	SDA	-	-
Automation (P3)						
Resourced required	FabSim, CARTO map construction platform	-	UNHCR, ACLED, Population databases (like Citypop), Bing Maps	SDA	Conference and project funding (Science Hackathon Geneva 2017 by CERN, Collaborations Workshop by SSI, ComPat project	Team, Academic
Resources created	FabFlee	-	-	-	-	-
Gamification (P4)						
Resources required	Kivy	-	UNHCR, geospatial	RFLP	-	Team, Academic
Resources created	-	-	-	-	-	-
2018						
Parallelisation and multiscale computing (P5)						
Resources required	MPI4Py	-	-	-	ComPat project	Single developer (DG)
Resources created	Parallelised and Coupling implementation files (pflee.py; (coupling.py)	-	-	Multiscale simulation approach	-	-
Input data feasibility (P6)						
Resources required	Tableau, Excel	-	UNHCR, ACLED	Correlation between dependent (refugee count) and independent variables (type of conflict event, magnitude of event, fatalities)	ComPat project	Team, Academic
Resources created	-	-	-	-	-	-
2019						
Coupling – Food security sub-model (P7)						
Resource required	FabFlee	Secondary sources	UNHCR, ACLED, Bing Maps , IPC	SDA, Temporal and spatial correlation	HiDALGO and VECMA	Team, Academic

Resource created	FabFlee (new commands)	-	-	-	-	-
Coupling – conflict evolution sub-model (P8)						
Resource required	Eagle SC (PSNC) (HiDALGO), FabFlee	Medicins sans Frontiers (on terminology), secondary sources	UNHCR, Geospatial, ACLED, MSCI World Metals and Mining Index	Hybrid simulation modelling, SDA	HiDALGO and VECMA	Team (HA joined), Academic
Resource created	-	-	-	-	-	-
2020						
Automated ensemble simulations (P9)						
Resources required	FabFlee	Secondary	UNHCR, ACLED, Geospatial	SDA	HiDALGO and VECMA	Team
Resources created	-	-	-	Automated policy exploration toolkit with integrated sensitivity analysis, extended SDA for counterfactual scenarios	-	-
2021						
Multiscale migration modelling (P10)						
Resources required	MUSCLE 3, Eagle supercomputer	Secondary	ECMWF Climate Data Store, GloFAS Data (Copernicus project), UNHCR, ACLED, Geospatial	SDA, Multiscale simulation approach, File I/O, Acyclic (one-way coupling), hybrid simulation approach	HiDALGO	Team (AJ joined), HiDALGO, Expertise on MUSCLE3 (Lourens Veen from the Dutch e-Science Centre)
Resources created	-	-	-	-	-	-
Scaling parallelised Flee (P11)						
Resources required	Hawk, Eagle, Vulcan supercomputers, pflee (python dependencies, numpy; scipy, MPI4Py)	-	UNHCR, ACLED, Geospatial	Basic agent parallelisation, agent space parallelisation	HiDALGO	Team, HiDALGO
Resources created	-	-	-	-	-	-
Automated location graph construction (P12)						
Resources required	Rule based AI, Open Source Routing Machine from Open Street Maps (OSM), HPE Apollo SC (not within HiDALGO consortium)	-	Open Street Maps	Triangle inequality, contraction hierarchies algorithm	HiDALGO, Austrian COMET Program (KNOW Centre)	Team, HiDALGO
Resources created	Route pruning algorithm					
Formalisation of Flee 2.0 rule set (P13)						
Resources required	VECMAtk (EasyVVUQ, QCG-PilotJob), FabFlee	IOM and NGOs on the field	UNHCR, ACLED, Geospatial	Sensitivity analysis (Stochastic collocation and Sobol sensitivity indices), SDA	VECMA and HiDALGO	Team and VECMA
Resources created	Flee rule set 2.0	-	-	SDSD	-	-
Multiobjective optimisation (P14)						
Resources required	SDA, Evolutionary algorithm (NSGA-II), FabFlee, Route pruning algorithm	-	UNHCR, ACLED, Geospatial	Facility Location Problem (FLP)	ITFLOWS, HiDALGO	Team, Academic, HiDALGO
Resources created	-	-	-	Multiobjective simulation	-	-

				optimisation approach		
Integrated refugee IDP forecasting (P15)						
Resources required	Flee	Qualitative reports, News (Associated Press, Reuters, BBC); Feedback from Save the Children	UNHCR, ACLED, OpenStreet Maps	SDA	HiDALGO	Team, Save the Children
Resources created	-	-	-	-	-	-

9.3 Appendix 3: Bibliography (Chapter 5 Case description and findings)

Anastasiadis, P., Gogolenko, S., Papadopoulou, N., Lawenda, M., Arabnejad, H., Jahani, A., & Mahmood, I., & Groen, D. (2021). P-FLEE: An efficient parallel algorithm for simulating human migration. *2021 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW),* 1008-1011. https://doi.org/10.1109/IPDPSW52791.2021.00159.

Berman, N., Couttenier, M., Rohner, D., & Thoenig, M. (2017). This Mine is Mine! How minerals fuel conflicts in Africa. *American Economic Review, 107*(6),1564-1610.

Boesjes, F. (2022). *Forced displacement in Mali Analysing the effect of physical environment data on refugee flight routes in Flee*. Masters thesis: University of Utrecht. https://studenttheses.uu.nl/handle/20.500.12932/41541.

Boesjes, F., Groen, D., Ooink, B., & Jahani, A. (2022). Forced displacement in Mali - Analysing the effect of the physical environment on refugee flight routes. *ICCS 2022 Presentation.* https://easychair.org/smart-program/ICCS2022/2022-06-22.html#talk:193427.

Campos, C.V. Suleimenova, D., & Groen, D. (2019). A coupled food security and refugee movement model for the South Sudan conflict. *Computational Science - ICCS 2019 Conference Paper, Lecture notes in Computer Science book series (LNTCS vol. 11540).*

Chan, N.T., Suleimenova, D., Bells, D. & Groen, D. (2018). Modelling refugee escaping violent events: A feasibility study from an input data perspective. In A. Anagnostou, R. Meskarian, D. Robertson, & M. Fakhimi (Eds). *Proceedings of the Operational Research Society Simulation Workshop 2018 (SW18),* pp. 156-163.

Clayton, J. (2015). *UNHCR chief issues key guidelines for dealing with Europe's refugee crisis.* United Nations High Commissioner for Refugees. https://www.unhcr.org/news/stories/unhcr-chief-issues-key-guidelines-dealing-europes-refugee-crisis. [01 August 2023].

Estrada, L. E. P., Groen, D., & Ramirez-Marquez, J. E. (2017). A serious video game to support decision making on refugee aid deployment policy. Procedia Computer Science, 108, 205–214. https://doi.org/10.1016/j.procs.2017.05.112.

European Union. (n.d.). *The EU's open science policy*. https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en [01 August 2023].

Groen, D. (2016). Simulating refugee movements: Where would you go? *Procedia Computer Science, 80*. https://doi.org/10.1016/j.procs.2016.05.400.

Groen, D., Bhati, A.P., Suter, J., Heatherington, J., Zasada, S.J., & Coveney, P.V. (2016). FabSim: Facilitating computational research through automation on large-scale and distributed e-infrastructures. *Computer Physics Communications, 207*(2016), 375-385. http://dx.doi.org/10.1016/j.cpc.2016.05.020.

Groen, D. (2018). Development of a multiscale simulation approach for forced migration. Computational Science – ICCS 2018. ICCS 2018. Lecture Notes in Computer Science – Vol. 10861. https://doi.org/10.1007/978-3-319-93701-4_69.

Groen, D., Bell, D., Arabnejad, H., Suleimenova, D., Taylor, S.J.E, & Anagnostou, A. (2019a). Towards modelling the effects of evolving violence on forced migration. *Proceedings of the 2019 Winter Simulation Conference*, 297-307.

Groen, D., Knap, J., Neumann, P., Suleimenova, D., Veen, L., & Leiter, K. (2019b). Mastering the scales: a survey on the benefits of multiscale computing software. Philosophical Transactions of the Royal Society A, *377*. http://dx.doi.org/10.1098/rsta.2018.0147.

Groen, D., Arabnejad, H., Jancauskas, V., Edeling, W. N., Jansson, F., Richardson, R. A., Lakhlili, J., Veen, L., Bosak, B., Kopta, P., Wright, D. W., Monnier, N., Karlshoefer, P., Suleimenova, D., Sinclair, R., Vassaux, M., Nikishova, A., Bieniek, M., Luk, Onnie O., Kulczewski, M., Raffin, E., Crommelin, D., Hoenen, O., Coster, D. P., Piontek, T., & Coveney, P. V. (2021). VECMAtk: a scalable verification, validation and uncertainty quantification toolkit for scientific simulations. Philosophical Transactions of the Royal Society A, *379*(20200221), 1-22. https://doi.org/10.1098/rsta.2020.0221.

Guillén, A. & Teodoro, E. (2023). Embedding ethical principles into AI predictive tools for migration management in humanitarian action. *Social Sciences, 12*(53). https://www.mdpi.com/2076-0760/12/2/53.

HiDALGO. (2019a). *D3.1 Report on benchmarking and optimisation.* https://hidalgo-project.eu/reports/wp3 [20 June 2022].

HiDALGO. (2019b). *D3.2 Initial specifications for HPC scalability optimisation, HPDA model implementation, data management, visualisation and coupling technologies.* https://hidalgo-project.eu/reports/wp3 [20 June 2022].

HiDALGO. (2019c). *D4.1 Initial status of pilot applications*. https://hidalgo-project.eu/reports/wp4 [20 June 2022].

HiDALGO. (2021a). *D3.5 Final report on benchmarking, implementation, optimisation strategies and coupling technologies.* https://hidalgo-project.eu/reports/wp3 [20 June 2022].

HiDALGO. (2021b). *D4.4 Final implementation report of the pilot and future applications.* https://hidalgo-project.eu/reports/wp4 [20 June 2022].

HiDALGO. (2021c). *D6.6 Final report on requirements, components and workflow integration* https://hidalgo-project.eu/reports/wp6 [20 June 2022].

ICRC. (2022). *Climate change in Mali: "We drilled deep but found nothing".* International Committee of the Red Cross. Retrieved from https://www.icrc.org/en/document/climate-change-mali-we-drilled-deep-found-nothing [20 July 2023].

IFRC. (2020). *World Disasters Report 2020.* International Federation of Red Cross and Red Crescent Societies. https://media.ifrc.org/ifrc/world-disaster-report-2020.

ITFLOWS (2021a). *D2.1 Report on the ITFLOWS international and European legal frameworks on migrants and refugees and ITFLOWS ethical framework*. https://www.itflows.eu/document-downloads/#deliverables.

ITFLOWS (2021b). *D2.3 Report on human rights, ethical, societal and data protection risks assessments*. https://www.itflows.eu/document-downloads/#deliverables.

ITFLOWS (2021c). *D2.4 Report on the ITFLOWS regulatory model*. https://www.itflows.eu/document-downloads/#deliverables.

ITFLOWS. (2021d). Analysis on migration drivers and trajectories along the eastern Mediterranean route (South-Central Asia & Middle East); Central and western Mediterranean Routes and the Western Africa Route (North, West, and the Horn of Africa), and the Atlantic Route (South-Central America). https://www.itflows.eu/document-downloads/#deliverables.

ITFLOWS. (2021e). *D7.1 Report on Users' Board participatory feedback.* https://www.itflows.eu/document-downloads/#deliverables.

ITFLOWS. (2022). *D3.4 Time sequence of forced displacement into neighbouring countries.* https://www.itflows.eu/document-downloads/#deliverables.

Jahani, A., Arabnejad, H., Suleimenova, D., Vuckovic, M., Mahmood, I., & Groen, D. (2021). Towards a coupled migration and weather simulation: South Sudan conflict. *Computational Science - ICCS 2021 Conference Paper, Lecture notes in Computer Science book series (LNTCS vol. 12746).*

Lowcock, M. (2019). *Anticipation saves lives: How data and innovative financing can help improve the world's response to humanitarian crises*. Speech delivered at the London School of Economics. London, 2 December. https://reliefweb.int/report/world/mark-lowcock-under-secretary-general-humanitarian-affairs-and-emergency-relief.

PwC. (2017). *Managing the refugee and migrant crisis: The role of governments, private sector and technology*. PwC Global Crisis Centre.

Redfern, C. (2022). *How the focus on Ukraine is hurting other humanitarian responses.* The New Humanitarian. https://www.thenewhumanitarian.org/news-feature/2022/07/07/Ukraine-aid-Russia-invasion-funding-donors [31 July 2023].

Schweimer, C., Geiger, B.C., Wang, M., Gogolenko, S., Mahmood, I., Jahani, A., Suleimenova, D., & Groen, D. (2021). A route pruning algorithm for an automated geographic location graph construction. *Scientific Reports, **11***(11547), https://doi.org/10.1038/s41598-021-90943-8.

Spindler, W. (2015). *2015: The year of Europe's refugee crisis.* United Nations High Commissioner for Refugees. https://www.unhcr.org/news/stories/2015-year-europes-refugee-crisis [01 August 2023].

Suleimenova, D. (2020). *Predicting forced displacement using a generalised and automated agent-based simulation.* PhD thesis: Brunel University London. https://bura.brunel.ac.uk/handle/2438/20732.

Suleimenova, D. & Groen, D. (2020). How policy decisions affect refugee journeys in South Sudan: A study using automated ensemble simulations. *Journal of Artificial Societies and Social Simulation, 23*(1), https://doi.org/10.18564/jasss.4193.

Suleimenova, D., Arabnejad, H., Edeling, W.N., & Groen, D. (2021). Sensitivity driven simulation development: A case study in forced migration. *Philosophical Transactions of the Royal Society A, 379*(20200077), 1-18. https://doi.org/10.1098/rsta.2020.0077.

Suleimenova, D., Bell, D., & Groen, D. (2017a). A generalised simulation development approach for predicting refugee destinations. *Scientific Reports, 7*(13377), 1-13. https://doi.org/10.1038/s41598-017-13828-9.

Suleimenova, D., Bell, D., & Groen, D.. (2017b). Towards and automated framework for agent-based simulation of refugee movements. *Proceedings of the 2017 Winter Simulation Conference*, 1240-1251. https://doi.org/10.1109/WSC.2017.8247870.

Suleimenova, D., Low, W., & Groen, D. (2022). An agent-based forced displacement simulation: A case study of the Tigray crisis. *Computational Science - ICCS 2022 Conference Paper, Lecture notes in Computer Science book series (LNTCS vol.* 13353*).*

UN. (2021). *UN 2.0 Quintet of Change.* https://www.un.org/two-zero/en [31 July 2023].

UNDRR. (2015). *Sendai framework for disaster risk reduction 2015-2030.* United Nations Office for Disaster Risk Reduction. https://www.undrr.org/implementing-sendai-framework/what-sendai-framework.

UNHCR. (2019). *Data transformation strategy: 2020-2025: Supporting protection and solutions*. United Nations High Commissioner for Refugees. https://www.unhcr.org/media/data-transformation-strategy-2020-2025 [01 August 2023].

UNOCHA FTS. (2021). *COVID-19 Global Humanitarian Response Plan.* United Nations Office for the Coordination of Humanitarian Affairs Financial Tracking Service.

UNOCHA. (2018). *World humanitarian data and trends.* United Nations Office for the Coordination of Humanitarian Affairs. https://reliefweb.int/report/world/world-humanitarian-data-and-trends-2018.

UNOCHA. (2020). *The Global Humanitarian Overview 2021.* United Nations Office for the Coordination of Humanitarian Affairs.

UNOCHA. (2021). *From digital promise to frontline practice: New and emerging technologies in humanitarian action*. United Nations Office for the Coordination of Humanitarian Affairs. https://reliefweb.int/report/world/digital-promise-frontline-practice-new-and-emerging-technologies-humanitarian-action.

UNOCHA. (2023). *The state of open humanitarian data 2023: Assessing data availability across humanitarian crisis*. UNOCHA Centre for Humanitarian Data. https://reliefweb.int/report/world/state-open-humanitarian-data-2023-assessing-data-availability-across-humanitarian-crises [01 August 2023].

VECMA. (2019). *D4.2: Report on the implementation of non- intrusive VVUQ techniques.* https://www.vecma.eu/about/deliverables/ [15 July 2023].

VECMA. (2021a). *D4.3: Report on the implementation of non- intrusive and intrusive VVUQ techniques.* https://www.vecma.eu/about/deliverables/ [15 July 2023].

VECMA. (2021b). *D5.3: Final report on the VECMA infrastructure.* https://www.vecma.eu/about/deliverables/ [15 July 2023].

Xue, Y., Li, M., Arabnejad, H., Suleimenova, D., Jahani, A., Geiger, B.C., Zidong, W., Liu, X., & Groen, D. (2022). Camp location selection in humanitarian logistics: A multiobjective simulation optimization approach. *Computational Science - ICCS 2022 Conference Paper, Lecture notes in Computer Science book series (LNTCS vol.* 13352*).*

9.4 Appendix 4: Corpus

Meetings and dates	
Team meetings	#1(15/07/2021); #2(20/07/2021); #3(27/07/2021); #4(29/07/2021); #5(05/08/2021); #6(07/09/2021); #7(13/09/2021); #8(17/09/2021); #9(20/09/2021); #10(23/09/2021); #11(27/09/2021); #12(30/09/2021); #13(04/10/2021); #14(07/10/2021); #15(11/10/2021); #16(14/10/2021); #17(18/10/2021); #18(21/10/2021); #19(25/10/2021); #20(01/11/2021); #21(04/11/2021); #22(08/11/2021); #23(15/11/2021); #24(18/11/2021); #25(22/11/2021); #26(29/11/2021); #27(07/12/2021); #28(16/12/2021); #29(10/01/2022); #30(13/01/2022); #31(18/01/2022); #32(24/01/2022); #33(03/02/2022); #34(07/02/2022); #35(17/02/2022); #36(25/02/2022); #37(03/03/2022); #38(10/03/2022); #39(17/03/2022); #40(21/03/2022); #41(31/03/2022); #42(05/04/2022); #43(21/04/2022); #44(26/04/2022); #45(05/05/2022); #46(11/05/2022); #47(19/05/2022); #48(24/05/2022); #49(09/06/2022); #50(16/06/2022).
External partners	
Humanitarian organisations	#1(04/05/2021); #2(02/08/2021); #3(07/10/2021); #4(12/11/2021); #5(05/04/2022); #6(08/06/2022).
Others	#1(30/07/2021); #2(06/09/2021); #3(06/10/2021); #4(16/11/2021); #5(18/11/2021); #6(28/01/2022); #7(25/05/2021); #8(08/06/2022).
Project meetings	
HiDALGO	#1(25/08/2021); #2(07/09/2021); #3(08/09/2021); #4(13/09/2021); #5(17/09/2021); #6(29/09/2021); #7(06/10/2021); #8(08/10/2021); #9(19/10/2021); #10(20/10/2021); #11(02/11/2021); #12(03/11/2021); #13(17/11/2021); #14(07/12/2021); #15(14/12/2021); #16(12/01/2022); #17(18/01/2022); #18(25/01/2022); #19(09/02/2022); #20(22/02/2022); #21(01/03/2022); #22(21/04/2022); #23(27/04/2022); #24(27/04/2022) [#23 and #24: separate meetings]; #25(28/04/2022).
VECMA	#1(16/07/2021); #2(26/07/2021); #3(06/09/2021); #4(20/09/2021); #5(04/10/2021); #6(18/10/2021); #7(19/10/2021); #8(19/10/2021) [#7 and #8: separate meetings]; #9(01/11/2021); #10(15/11/2021); #11(10/01/2022).
ITFLOWS	#1(16/09/2021); #2(16/09/2021) [#1 and #2: separate meetings]; #3(06/10/2021); #4(22/10/2021); #5(16/11/2021); #6(26/11/2021); #7(03/12/2021); #8(10/12/2021); #9(15/12/2021); #10(11/01/2022); #11(08/02/2022); #12(23/02/2022); #13(09/03/2022); #14(23/03/2022); #15(05/04/2022); #16(03/05/2022); #17(04/05/2022); #18(18/05/2022); #19(01/06/2022).
SEAVEA	#1(02/09/2021); #2(08/12/2021); #3(17/01/2022); #4(24/01/2022); #5(21/02/2022); #6(07/03/2022); #7(21/03/2022); #8(04/04/2022).
Workshops	
Workshops	#1(23/07/2021: VECMAtk training workshop); #2(24/09/2021: Flee workshop); #3(29/09/2021: ITFLOWS public webinar); #4(17/11/2021: YX's MOO seminar at BUL); #5(21/01/2022: Predictive modelling of forced displacement workshop by UN Global Pulse, UNHCR Innovation; UNHCR Brazil Boa Vista sub-office); #6(29/04/2022: DG's lecture on Flee at Oxford Brookes); #7(30/11/2022: KEI Event convened by author with support from LSE Knowledge Exchange and Impact Fund in partnership with the team at BUL).
Documents	
Internal documents	Tigray conflict (reports, works-in-progress, assumptions) x 14; Rule set x 2; Columbia University requirement specifications x 1
Publications and project deliverables	
Journal articles, conference papers, dissertation, PhD thesis	#1(Groen et al., 2016); #2(Groen, 2016); #3(Suleimenova et al., 2017a including Supplement); #4(Suleimenonva et al., 2017b); #5(Estrada et al., 2017); #6(Groen, 2018); #7(Chan et al., 2018); #8(Campos et al., 2019); #9(Groen et al., 2019a); #10(Groen et al., 2019b); #11(Suleimenova, 2020); #12(Suleimenova & Groen, 2020); #13(Jahani et al., 2021); #14(Anastasiadis et al., 2021); #15(Schweimer et al., 2021); #16(Suleimenova et al., 2021); #17(Groen et al., 2021); #18(Suleimenova et al., 2022), #19(Xue et al., 2022); #20(Boesjes, 2022); #21(Boesjes et al., 2022).
HiDALGO project reports	#1(HiDALGO, 2019a); #2(HiDALGO, 2019b), #3(HiDALGO, 2019c); #4(HiDALGO, 2021a), #5(HiDALGO, 2021b); #6(HiDALGO, 2021c); #7(HiDALGO, 2022)*
VECMA project reports	#1(VECMA, 2019, D4.1)*, #2(VECMA, 2019); #3(VECMA, 2019, 5.1)*; #4(VECMA, 2021a); #5(VECMA, 2021, 4.4)*; #6(VECMA, 2021b).
ITFLOWS project reports	#1(ITFLOWS, 2021, D9.3)*; #2(ITFLOWS, 2021, 3.1)*; #3(ITFLOWS, 2021d); #4(ITFLOWS, 2021, D4.3)*; #4(ITFLOWS, 2021, D6.1)*; #5(ITFLOWS, 2021b); #6(ITFLOWS, 2021c); #7(ITFLOWS, 2021, Ethics Handbook)*; #8(ITFLOWS, 2021a); #9(ITFLOWS, 2021e); #10(ITFLOWS, 2022, D4.2); #11(ITFLOWS, 2022, D6.2)*; #12(ITFLOWS, 2022, Interim project report)*; #13(ITFLOWS, 2022); #14(ITFLOWS, 2022, D9.4)*; #15(ITFLOWS, D8.1)*.
Preliminary interviews	
Preliminary interviews	#1(DS, 15/07/2021); #2(DG, 15/07/2021); #3(HA, 06/08/2021); #4(DG, 18/10/2021), #5(DS, 21/10/2021).

* Part of the corpus but was not included in the process narrative in Chapter 5. Projects deliverables were prepared in successive iterations according to milestones in funding agreements, for example, D4.1, 4.2 etc. D4.1 would first milestone report of work package 4 and so on. The convention was the same across projects. Not all documents in the corpus were cited in the narrative due to overlaps in information, information better covered in deliverables cited, and to ensure concise focus on the case. Nevertheless, they contributed to an overall understanding of the process.