Grant Agreement Number: 825225

Safe-DEED

www.safe-deed.eu

D2.1 Threat and Incentive Model

Deliverable number	*D2.1*
Dissemination level	*Public*
Delivery date	*29 November 2019*
Status	*Final*
Author(s)	*Jeevan Jeevan Kumar Narayana Swamy* *Mark de Reuver* *Wirawan Agahari* *Tobias Fiebig*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825225.

This page has been intentionally left blank.

Changes Summary

Date	Author	Summary	Version
08.08.2019	Jeevan Swamy	First draft	0.1
10.09.2019	Tobias Fiebig	Second draft	0.2
15.10.2019	Tobias Fiebig	Third draft for review by Mark de Reuver	0.3
17.10.2019	Mark de Reuver	Revision	0.4
22.10.2019	Wirawan Agahari	Revision on Chapter 2	0.5
23.10.2019	Tobias Fiebig	Revisions of review by Mark de Reuver and Wirawan Agahari	0.6
24.10.2019	Tobias Fiebig	Version for internal review	0.7
18.11.2019	Tobias Fiebig	Incorporation of review feedback	1.0

This page has been intentionally left blank.

Executive summary

The emergence of data sharing to fuel business innovation is the latest iteration in the phenomenon of the data-driven transformation of the world. Data marketplaces have emerged as a new form of data-driven business models which enable trading of data between the data owners/providers and data consumers by providing the necessary technological and non-technological infrastructure. However, so far we have not seen a widespread adoption of data marketplaces. Safe-DEED aims at changing this, enabling the EU economic area to move towards a data driven economy. To ensure that this transformation is secure, the use of privacy preserving technologies is imperative. One of the most common privacy enhancing technologies for data sharing is secure Multi-Party Computation (MPC) technology, which provides a solution to these problems. Through its capabilities to preserve the confidentiality of data architecturally and thereby securing the interests of the data actors with respect to the uncertainty of the threat landscape around data, MPC can enable safe and secure data sharing between data actors. This characteristic of MPC can help data marketplaces to overcome their challenges and foster their realization. However, we still lack an understanding of (a) how MPC changes the threat landscape of data marketplaces, and (b) what incentives companies have to adopt MPC technology, given that it comes with several challenges.

We investigate these two questions in this report by constructing four conceptual models. The first two models are Pre-MPC Data Marketplace Platforms (a high-level architecture of the data marketplace platform) and Post-MPC Data Marketplace Platform (MPC incorporated architecture). We use the difference of these two models to understand the implications of MPC on the architectural aspects of data marketplaces. The other two models comprised of Pre-MPC Threat Model (threats to data marketplaces prior to MPC incorporation) and Post-MPC Threat Model (threats after MPC incorporation). We use these two models to study the implications of MPC on the threat landscape of data marketplaces.

The development of the four conceptual models was carried out in two phases. In each case, we constructed an initial model based on an structured literature survey. We then validate these models using expert interviews.

We find that MPC technology eliminates the business threats of ***Loss of Control over Data, Data Leakage, Data Leakage by Back Correlation, Loss of Competitive Advantage for Data Actors, Regulatory Threats*** and ***Data Sensitivity.*** Hence, MPC eliminates serious business threats associated with the issue of ***Data Sensitivity*** in a ***Security-by-Design*** way. This, in turn, reduces the burden of the ***Governance Model*** on its technological front. Hence, Incentive structures for the adoption of MPC in data marketplaces should focus on the operational benefit of a streamlined governance model. In our research, we find that the potentially negative implications of MPC technology thereby still not outweigh the positive synergies created by the reduced need for interparty trust and limited exposure of data assets. The deliverable thus affirms the idea that privacy and confidentiality preserving technologies, such as MPC, increase the incentives for sharing data by both enhancing trust and reducing security risks.

This page has been intentionally left blank.

Table of Contents

List of Figures

This page has been intentionally left blank.

List of Tables

This page has been intentionally left blank.

Abbreviations

ATT&CK	Adversarial Tactics, Techniques & Common Knowledge Framework
B2B	Business-to-Business
CAPEC	Common Attack Pattern Enumeration and Classification
CIA	Confidentiality, Integrity, Availability
COI	Community-of-Interest
CTSA	Cyber Threat Susceptibility Analysis
(D)DoS	(Distributed) Denial of Service
DevOps	Software Development (*Dev*) and Information-Technology Operations (*Ops*)
DL4LD	Data Logistics for Logistics Data
DLT	Distributed Ledger Technology
DMA	Data Market Austria
DMP	Data Marketplace Platform
DREAD	Damage, Reliability, Exploitability, Affected Users, Discoverability
DVT.	Data Valuation Technology
EDM.	Enterprise Data Market
FFIEC.	Federal Financial Institutions Examination Council
FSS	Financial Services Sector
GDPR.	General Data Protection Regulation
HLA.	High-Level Architecture framework
HLCTM.	High-Level Cyber Threat Modelling Framework
HSSEDI.	Homeland Security Systems Engineering & Development Institute
IDDIL/ATC	Identify the assets; Define the attack surface; Decompose the system; Identify attack vectors; List threat actors; Analysis & assessment; Triage; Controls.
Intel's TARA.	Threat Agent Risk Assessment
IoT	Internet of Things
Microsoft SDL	Security Development Lifecycle
MITRE'S TARA	Threat Assessment and Remediation Analysis

MPC	Multi-Party Computation
NGCI	Next Generation Cyber Infrastructure
NIST	National Institute of Standards and Technology
OWASP	Open Web Application Security Project
PII	Personal Identifiable Information
PPT	Privacy=Preserving Technologies
SaaS	Software as a Service
Safe-DEED	Safe Data Enabled Economic Development
SDLC	Systems/Software Development Life Cycle
SESAR	Single European Sky ATM(Air Traffic Management) Research
SSH	Secure SHell encryption protocol
STRIDE	Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege
TTP	Tactics, Techniques, Procedures

1 Introduction

This deliverable reports on the results of Task 2.1 in WP2 of the Safe-DEED*: Safe Data Enabled Economic Development* project (Lupu, 2018). It is a consortium of research organizations from cryptography, data science, business model innovation and legal domains across Europe, which is funded by the *European Union's Horizon 2020 Research and Innovation* program, to foster and accelerate the data-driven economy in Europe.

Over the past decade, data emerged to be one of the most valuable business resources disrupting and fueling the transformation of industries, leading to the statement, "Data is the new oil" (Hartmann, Zaki, Feldmann, & Neely, 2016). This so-called 'Big Data Revolution' encouraged businesses to adopt data-driven innovation which could potentially improve their productivity and efficiency. The knowledge extracted from the data and in many cases, the data itself have helped organizations to create enormous value in the form of data-driven decision making and data-based products respectively (Davenport, 2006; Brynjolfsson & McAffee, 2012). The value of data economy fostered by the activities of generation, collection, storage, processing, distribution, analysis and exploitation of data is expected to be around €700 Billion by 2025 in Europe alone (Lupu, 2018). To achieve this forecast, a well-coordinated interaction between demand and supply of the data is necessary. Data Marketplaces play a fundamental role in orchestrating this interaction by offering a platform equipped with different services for data owners to sell their data and for data seekers to find good quality data of their interest (Koutroumpis, Leiponen, & Thomas, 2017; Deichmann, Heineke, Reinbacher, Wee, 2016). Through simplifying data supply chains by overcoming the cumbersome logistics currently involved in searching, buying and selling data, data marketplaces help in establishing data ecosystems comprising of a network of organizations across different industries, and thereby could boost the data-driven economy.

However, despite their significance, the number of successful commercial data marketplaces are surprisingly low, and the number of failed ones is very high (Koutroumpis et al., 2017). The reason for this phenomenon can be twofold.

1. An intrinsic reason that it is challenging to design and set up a technologically viable platform to trade data (Koutroumpis et al., 2017). The immaturity and unavailability of enabling technologies presents a difficulty in developing a sound platform for a data marketplace.
2. An external reason associated with the lack of trust among data actors which is manifested by the uncertainty associated with data security due to the sensitive nature of data (Lupu, 2018). For data-driven innovation to flourish, it is crucial that the data owners share their data. However, several issues discourage them to do so. Some of the issues are: lack of clarity in the implementation of *General Data Protection Regulation* (GDPR), risks associated with privacy violations, threats associated with the business and cyber space around the data marketplaces. Fully concerned with these issues, data actors exhibit reluctance towards participating in data marketplaces. Since the data marketplaces are platforms prone to positive externalities, this reluctance of data actors directly implies a barrier which significantly contributes towards their slow and delayed flourishing of data marketplaces.

Multi-Party Computation (MPC) technology, provides a solution to these issues with their capabilities to safeguard the data and thereby, the interests of the data actors (Lupu, 2018).

Safe-DEED proposes to develop technologies to promote the data sharing culture among the organizations and foster the data-driven economy in Europe. The technologies proposed by Safe-DEED are of two categories*: Secure Multi-Party Computation* (MPC) and *Data Valuation Technologies* (DVT). With these technologies, Safe-DEED aims to encourage data owners to share their data by taking care of two crucial things:

- Enforcing the security aspects of the data sharing (through Secure MPC); and
- Explicating the value of the data held by the data owners to both data owners and data consumers (through DVT)

By ensuring these two aspects, Safe-DEED aims to incentivize data actors to share more data, thereby enabling a data driven economy in Europe.

1.1 T2.1 Task Description

Several researchers proposed Multi-Party Computation (MPC) technology as a crucial enabler of the data marketplaces which has the potential to orchestrate safe and secure data trading (Roman & Stefano, 2016; Lupu, 2018). However, to ensure that these technologies are adopted, we have to understand (a) Which threats they mitigate and possibly introduce, and (b) what incentives for companies exist to adopt MPC. Hence in this task, we will construct and validate incentive models for the adoption of privacy and confidentiality preserving technologies. We will generate these models from a multi-actor perspective, including users' perspective as well as the perspective of organizations which have to procure privacy and computer security preserving technologies.
Based on the feedback of our industry partners, we understood that the major driver for companies' adoption of secure MPC will be based on the impact this technology has on the threat landscape. Hence, the main objective of this task converges in the question of how MPC *changes* the threat landscape for data sharing in data marketplaces. While the initial Safe-DEED proposal foresaw that the threat model would be validated through WP6 and WP7 use cases, this turned out not to be feasible, since no technical architecture for data sharing had been defined and implemented at the time of writing the deliverable. Further, while the Description of Action suggests secondary cases, in reality there are no implementations of MPC available in practice. For these reasons, we rely on a generalized architecture and threat model (the HLA and HLTM respectively). An advantage of doing so is also that our findings extend beyond the specific use cases in WP6 and WP7 and are a valid starting point for understanding the threats and incentives of data sharing in any context, within or outside the scope of Safe-DEED.

1.2 Knowledge Gap

Firstly, related to the *architectural* aspects of the data marketplaces, there has been no investigation of how MPC technology can be incorporated into the data marketplace platform architecturally. Related to this, there exists another problem that there has been no research related to the architectural aspects of data marketplaces. The reasons for this can either be that the architectural information is confidential proprietary information for the real-life data marketplaces to disclose to the research community; or also that the research area of architectural aspects of the data marketplaces is in its infancy and is not explored proactively yet. As a result, there exists no architecture of a data marketplace platform in the literature. So, there is a need to build an architecture which reflects a generic data marketplace platform.
Secondly, it is necessary to investigate the effect of MPC technology on the *threat landscape* of the data marketplaces in order to understand the positive as well as negative implications of MPC technology towards the threats associated with the data marketplaces. Here exists another problem that the threats associated with the data marketplaces has never been identified yet by the research community. So, there is a need to explore the threat landscape of the data marketplaces to identify the threats that affect them.

1.3 Research Objective

We formulate the research objective (*RO*) for this task as follows: *"What are the implications of Multi-Party Computation (MPC) technology for the architecture and the threat landscape of Data Marketplaces that facilitate data sharing, and what incentives for the adoption of MPC exist?"*

Hence, in this task, we have to process the following steps:

1. Create a workable case, by developing a *generic* architecture of a data marketplace platform to be used in the threat and incentive modelling.
2. Explore *threat landscape* of data marketplaces to identify the threats which affect their functioning, i.e., perform threat modelling.
3. Understand how the MPC technology can be *incorporated* into the previously-built architecture to deduce what this implies architecturally for the data marketplaces
4. Deduce how MPC technology *affects* the threat landscape of data marketplaces (*both positive and negative effects*) and how this (de)incentivizes the adoption of MPC.

1.4 Research Design

We depict the research framework we created in **Figure 1**. We first conduct a structured literature review of the existing data marketplace and threat modelling literature. This feeds into the development of our business function and threat models. We then explore the integration of MPC into these models. All four models are the subjected to expert interviews for validation.

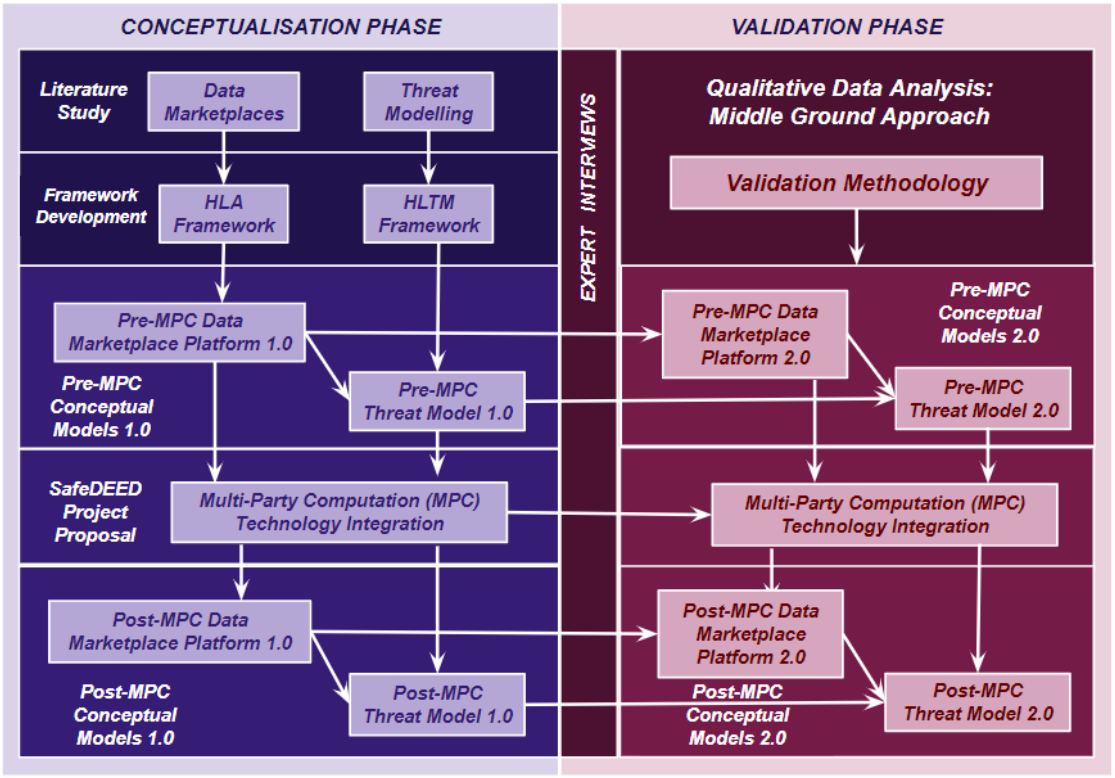

Figure 1: Task research framework

1.4.1 Conceptualization Design

We derive the conceptualization of the first iteration of the four conceptual models from a structured literature review. This phase was executed with the help of the following desk research methods:

- ***Structured Literature Study:*** Our Structured Literature Study (SLR) involved three common steps: *searching the literature, reviewing the selected literature and critically analyzing the obtained knowledge to appropriately use in building our conceptualizations*. Related to our research, the literature study was conducted extensively on the subjects; data marketplaces and threat modelling while additional research on the cyber threats associated with the information systems and MPC technology was conducted. To also reflect developments outside of academia, and to ensure a connection to practice, we did not only include academic sources (*journal articles, conference articles, theses*), but also non-academic ones (*consultancy articles, white papers, web articles et cetera*).
- ***Framework Development:*** Based on the results of our SLR, we develop the *HLA (High Level Architecture) framework* and *HLTM (High Level Threat Model) framework*, to illustrate the fundamental concepts associated with the technological entities in general and the cyber threat modelling of those technological entities respectively.
    - ***HLA Framework:*** We used the HLA framework to build the high-level architecture for a generic data marketplace platform which signified the *Pre-MPC Data Marketplace Platform 1.0.*
    - ***HLTM Framework:*** Similarly, this was used to build the high-level cyber threat model for the data marketplace platform consisting of the cyber threats associated with the business functions of the data marketplace platform.

1.4.2 Validation Design

Our validation activity involves refining, updated, modifying or invalidating the theoretical concepts develop based on the SLR. We perform a *qualitative study*, as qualitative data provides the flexibility needed to carry out exploration with inductive reasoning. The qualitative data was collected with the method of *interviews* given that it provides rich primary data about the phenomenon. The prospective participants for the interviews were selected to be *subject area experts*. Due to the cutting-edge non-mainstream nature of MPC for distributed data marketplaces, our sample pool was limited. As a result, j*udgement sampling* was carried out to scout for eligible experts (Sekaran & Bougie, 2013). The recruited experts comprise of *researchers* and *industry experts* in the subject areas of Data Marketplaces, Threat Modelling and MPC technology. The interviews were conducted in a cross sectional manner (Sekaran & Bougie, 2013).
We use a *Middle-Ground Approach* (Sekaran & Bougie, 2013) to analyze our research data, which is a variant of *Grounded Theory* (a very common method used to generate theoretical frameworks (Corbin & Strauss, 1990)).

1.5 Key Findings

We find that MPC technology eliminates the business threats of ***Loss of Control over Data, Data Leakage, Data Leakage by Back Correlation, Loss of Competitive Advantage for Data Actors, Regulatory Threats*** and ***Data Sensitivity.*** Hence, MPC eliminates serious business threats associated with the issue of ***Data Sensitivity*** in a ***Security-by-Design*** way. This, in turn, reduces the burden of the ***Governance Model*** on its technological front. Hence, Incentive structures for the adoption of MPC in data marketplaces should focus on the operational benefit of a streamlined governance model. The potentially negative implications of MPC technology—like an increased overhead in data processing complexity, or an inability to verify included data—thereby still not outweigh the positive synergies created by the reduced need for interparty trust and limited exposure of data assets.

Furthermore Task 2.1 of Workpackage 2 made the following contributions:

- A new taxonomy of data marketplace platform designs, which provides an updated classification comprising of the different platform designs containing both concept platforms and realized ones. The taxonomy refines the basic classification of Koutroumpis et al. (2017) and updates it with a variety of probable data marketplaces**.** This provides a new foundation to position different data marketplaces either during design or analysis.
- A new list of functional requirements was developed which furthers the conversation of the functional requirements from just being technological to also include non-technological aspects, helping to understand what is expected of data marketplaces.
- As a significant contribution to the gap in the literature involving the architectural aspects of the data marketplaces, this research presents the *High-Level Architecture* of a generic data marketplace platform**.** This architecture can act as a reference architecture for the researchers to build more sophisticated and detailed architectures for the data marketplace platforms. Additionally, the *HLA Framework 2.0* can be used by researchers to build high-level architectures for the technological entities.
- We develop a new *Business Threat Model* and a *Cyber Threat model* for a *generic* data marketplace platform. This threat models marks the first of its kind for data marketplaces.

Secondly, the task also contributes to the state of the art for *threat modelling***,** which mostly focuses on software centric threat modelling, lacking a business function focus:

- A new taxonomy for threat models was created to expand the scope of threat modelling from just the low-level cyber threats to also include high-level business threats. This taxonomy goes beyond just focusing on the cyberspace and includes the analyses of threats to the business logic of the focal entity. The NGCI Apex Classification of Cyber Threat Models by Bodeau et al. (2018) is also positioned in our taxonomy.
- A new cyber threat modelling framework which operates at the business function level of information systems of technological entities was developed which goes by the name *High-*

Level Cyber Threat Modelling (HLCTM) framework. This framework provides an effective threat model for detained architectures and provides a baseline threat model for high-level technological entities. Additionally, the framework provides a straight forward way to carry out low-profile threat modelling on technological entities which can be used for auxiliary tasks of researches in bigger scopes

Finally, the task also contributes to the gaps existing in the literature of the *MPC technology* mostly associated with its business application aspects. These are discussed as listed below:

- Our research clarified the business process of MPC technology finding that the process is dependent on the underlying use-case and hence, cannot be standardized for a platform like data marketplace. Instead, it can only be provisioned in an ad-hoc form. Furthermore, we explicated the application of this business process in a data marketplace platform. Thus, contributing an application for the gap involving the business application of MPC technology.
- Furthermore, we have also investigated the effect of MPC on the threats associated with data sensitivity and data marketplaces which furthers the literature explicating the advantages and shortcomings of MPC technology.

1.6 Structure

The remainder of this document is structured as follows. In Chapter 2, we introduce Background on Data Marketplaces, followed by the Background on Threat Modelling in Chapter 3. Next, we present Threat Modelling for Data Marketplaces in Chapter 4. We then outline the Implications of MPC on Data Marketplaces in Chapter 5 and perform our Model Validation in Chapter 6. Next, we provide a Discussion of our results in Chapter 7, and finally a Conclusion in Chapter 8.

This page has been intentionally left blank.

2 Background on Data Marketplaces

We conducted a literature study on data marketplaces to explore the phenomenon of data marketplaces and to understand their fundamental concepts. Based on our findings, we developed an architecture for a generic data marketplace platform, which we can subsequently use for threat modelling.

2.1 Literature Search and Selection Methodology

The focus of this literature study was to obtain an architecture of a generic data marketplace platform. With this objective in mind, we only included relevant literature that comprises the fundamental concepts associated with the data marketplaces. Examples include characteristics of data marketplaces, the underlying architecture, the functionalities and features, and actors within the ecosystems. These concepts are required as a basis to establish an architecture for a generic data marketplace platform.
To pursue this agenda, we performed a literature search on Scopus with the search phrase "***data marketplaces***" which yielded ***69*** articles. We only conduct the literature search until 10 May 2019, meaning that we did not consider any literature published after this date. We cluster our search results in four streams of literature on data marketplaces (see Table 1): (1) classification, (2) the economics perspective, (3) architectural aspects, and (4) general topics.

Table 1: Categories of current research on data marketplaces

Catagories	*Key references*	*Rationale*
Classification of data marketplaces	Muschalle, Stahl, Löser, & Vossen (2013); Schomm et al. (2013); Stahl, Schomm, & Vossen (2014); Stahl, Schomm, Vossen, & Vomfell (2016); Stahl, Schomm, Vomfell, & Vossen (2017)	Pioneers in the research of data marketplaces.
The economic perspective of data marketplaces	Koutroumpis & Leiponen, (2013); Koutroumpis, Leiponen, & Thomas (2017)	Provides basic concepts associated with the data marketplaces (e.g. business logic, challenges involved in setting up a data marketplace).
Architectural aspects of data marketplaces	Quix, Chakrabarti, Kleff, & Pullmann (2017); Chakrabarti, Quix, Geisler, Khromov, & Jarke (2018)	Relevant research on functionalities, features and actors of data marketplaces based on Industrial Data Spaces.
General topics on data marketplaces (e.g. big data, data commercialization, data contracts, metadata models)	Muschalle et al. (2013); Fricker & Maksimov (2017); Spiekerman et al. (2018)	Secondary and tertiary references that were identified via backwards and forward snowballing.

A review of the articles indicated that there is a scarcity of literature related to the basic functioning of data marketplaces. In particular, we found that several articles dealt with specific issues of data marketplaces such data pricing (Muschalle et al., 2013; Fricker & Maksimov, 2017) and metadata (Spiekermann et al., 2018), while majority of the articles proposed data marketplaces for specific domains (e.g. the automotive industry, health care industry, credit scoring). Few articles were mentioning the functionality aspect of the data marketplaces; however, it was not the main focus.
We found that the literature on the functioning of data marketplaces is still in its infancy. The reason can be that the initial focus was on figuring out how to price the data. Only recently, with the events of data marketplaces failing (Schomm, Stahl, & Vossen, 2013) or stopping their operations (Ramel, 2016), researchers got interested in investigating the functionality of data marketplaces. The advent of MPC and other enabling technologies crucial for the functioning of the data marketplaces also happened just recently, indicating the potential increase of interest in the future.

2.2 Data as a Commodity

To understand what makes the data marketplaces a unique species of business, we need to understand data as a trading commodity of the marketplaces. This knowledge is essential since data exhibit very different characteristics than a common good, which pose challenges for the successful commodification

of data. Building on the conceptual work by Koutroumpis et al. (2013; 2017), we discuss here two challenges: *Weak Protection Regime and Data Sharing Reluctance*.

2.2.1 Weak Protection Regime

Data has two characteristics that make it challenging to assign intellectual property (IP) rights to protect data effectively. First, data is *a non-rivalrous good* that can be replicated with negligible cost and can be used simultaneously at multiple locations by different entities (Koutroumpis et al., 2013). Second, data is an *intermediate good* which means that it is of less or no business value unless either subjected to analysis or when combined with data from other appropriate sources, thereby creating meaningful data products (Koutroumpis et al., 2017).
The copyright laws and the database rights protect data in the confines of a database, but neither the actual data contents nor their intangible knowledge (Koutroumpis et al., 2017). So, once the data is out of the database and is modified either subjecting to analysis or combining with other datasets, the rights do not apply to the resultant data content or the extracted knowledge; and hence, it becomes almost impossible to trace the path travelled by a data point (Koutroumpis et al., 2017) (also called as *Data Lineage*). This condition results in a *weak protection regime* which makes data a tricky commodity for trading.

2.2.2 Data Sharing Reluctance

Meanwhile, *data sharing reluctance* refers to conditions where data sellers are unwilling to sharing their data or may share low-quality data due to various concerns. According to Koutroumpis et al. (2013), data is an *experience good* where the buyer has less insight in the good than the seller, which leads to difficulties faced by sellers in judging the value of data. Also, data is suffering from *Arrow's paradox*, meaning that the value of data can only be assessed by disclosing it (Arrow, 1972). This paradox is resulted in a difficulty to transact high-value data. Furthermore, sellers are often unaware or being unclear about the legal status of the data due to the ignorance towards regulations like the General Data Protection Regulation (GDPR).

2.2.3 Implication of these Challenges

Both challenges imply that data needs to be coupled with the information about its *provenance* which includes its origin, history and properties. We refer this information as the "*metadata*" of the data good which helps in judging the credibility, quality and security status of the data. There has been considerable research on designing metadata models for data products (Koutroumpis et al., 2017; Spiekermann et al., 2018). At present, maintaining data provenance is difficult once the data is transacted. Therefore, data marketplaces should have a mechanism to address these challenges.

2.3 Overview of Data Marketplaces

In this section, we discuss the concepts acquired from the literature search on the topic of data marketplaces, starting from their definition, the issues involved in materializing them and their variants.

2.3.1 Definition of Data Marketplaces

Data marketplaces were first defined as platforms where registered data providers can upload and maintain datasets. In order to be able to access, manipulate and use the data, data consumers are granted access through varying licensing models (Schomm et al., 2013). The definition was based on a survey on data marketplaces. However, the inclusion criteria for the survey was inconsistent with the definition of platforms that brings together both sides of the data market.
Meanwhile, Deichmann et al. (2016) provided a more accurate definition of data marketplaces based on the context of Internet of Things (IoT) data. They define data marketplaces as: "*platforms that connect*

providers and consumers of datasets and data streams, ensuring high quality, consistency and security. The data suppliers authorize the marketplace to license their information on their behalf following defined terms and conditions". This definition holds suitable for any form of data as the focus is on the data being a commodity but not on its different types.

A more comprehensive overview of data marketplaces was provided by Koutroumpis et al. (2017). In their classification, only one category reflects the true platform version of the data marketplace where any data supplier can upload and sell data to any data consumer. They call this variant ***many-to-many*** or ***multilateral*** data marketplaces. We use the same definition for our research.

Consistent with Deichmann et al. (2016), Koutroumpis et al. (2017) defined *multilateral* data marketplaces as *multi-sided platforms* where a digital intermediary connects data sellers, data buyers and facilitates data sharing activities. This variant of data marketplaces only orchestrates the data exchange process through services of search/discovery, transaction validation, transaction history and payment gateway. Functionally speaking, multilateral data marketplaces enable the association of disparate datasets from different data owners through easy search and discovery, standardization of their formats and their subsequent aggregation into meaningful data products (Koutroumpis et al., 2017). This mandates the necessity of regulatory environment, communication standards, data protocols and procedures of data import, storage, transformation, aggregation, analysis and delivery functionalities (Koutroumpis et al., 2017).

With these services, like any other digital marketplace platform, data marketplaces create value for their customers in the following ways (Smith et al., 2016). Firstly, the search process for data is simplified without having to browse each data provider's offering at their websites. Secondly, access to rich data content is enabled, which can be compared with each other to make an informed decision. Thirdly, automated data exchange with standardized data formats makes the trading process more convenient. Finally, there is a broader scope for building relationships by an improved match between supply and demand of data.

Despite these advantages, only a few examples of functional data marketplace platforms exist. Recently, Microsoft Azure Data Marketplace, which was the first mover to establish data marketplace platform, closed its operations and transformed itself into a different marketplace providing sophisticated data products and analytics services; instead of just data. The reason for this was the lack of a customer base interested in using Microsoft Azure Data Marketplace (Ramel, 2016). Put differently, such data marketplaces experienced positive externalities which means that its value is decreasing if the number of participants decreases (Eisenmann, Parker, & Van Alstyne, 2006). Customers may not opting to join the data marketplace since the platform does not effectively address the challenges of commoditizing data, which results in a lesser trust to trade high-value commercial data. Hence, we can find many open data marketplaces in existence which offer data of lower value (*open data*); while a very small number of commercial data marketplaces.

2.3.1.1 Types of Data Marketplaces

Smith (2018) classifies data marketplaces into three categories based on the type of data and parties involved in the exchange of data, namely *personal data marketplace, business data marketplace and sensor data marketplace*. See Table 2 for details. Following the main objectives of Safe-DEED, we focus on *Business (B2B) Data Marketplaces*. Hence, we will perform modelling for a "***many-to-many B2B data marketplace***". For simplification, we refer to it as "*data marketplace(s)" or "data marketplace platform(s)"* in the rest of the deliverable.

Table 2: Types of data marketplaces (Smith, 2018)

Catagories	*Key references*	*Rationale*
Personal Data Marketplaces	A platform for individual consumers to monetize data on their terms to the concerned buyers.	Datum, DataWallet, Physical
Business Data Marketplaces	A platform that enables B2B data exchange by providing a standard data model and interface to trade data	Not yet implemented.
Sensor Data Marketplaces	A platform to trade real-time data streams from IoT sensors	IOTA DataMarket, DataBroker DAO, Steamr

2.3.2 Data Marketplace Platform Designs

In order to overcome the challenges of commoditizing data discussed in sub-section 2.2, Koutroumpis et al. (2017) propose three institutional requirements as listed below:

1. Strict ***boundary conditions*** to data marketplace platforms are instrumental in allowing only legitimate users to participate in the data transaction while filtering out unreliable users.
2. ***Rules of usage*** enable control over data for the data sellers through data contracts which specify the criteria for data usage, thus providing legal cover restricting the misuse of data.
3. ***Monitoring mechanism*** oversees all the data transactions and operations on the data marketplace platform and can detect any anomalous activity. This constitutes the governance aspect of the data marketplace platforms.

Building on these requirements, Koutroumpis et al. (2017) suggest three designs of *multilateral* data marketplace platforms (see Table 3 for an overview of these designs):

1. ***Centralized platforms,*** which holds data centrally and offer their services on a central technological platform. This platform design enforces strong *boundary conditions* through formal entrance policies but fails concerning *rules of usage* and *monitoring mechanism*.
2. ***Decentralized platform,*** which uses Distributed Ledger Technology (DLT) as a core technology to perform data sharing activities. This platform design enforces all institutional requirements but suffers from technological immaturity as the DLT is not scalable for large scale operations (Simonite, 2016). Nevertheless, there is a considerable amount of research going on to make this design a reality, such as Enigma (Zyskind, Nathan, & Pentland, 2015), Sterling (Hynes, Dao, Yan, Cheng, & Song, 2018), and Trusted Data Marketplace (Roman & Stefano, 2016).
3. ***Collective platform,*** which achieves the enforcement of institutional properties by forming a closed consortium of partners (*boundary conditions*) powered by complex contracts (*rules of usage*) and effective *monitoring mechanism* took care of *platform provider*. However, this design is only effective when they are formed by a small number of partners with pre-existing trust-based relationships and shared interests of data exchange.

Table 3: Typology of multilateral data marketplace, adapted from Koutroumpis et al. (2017)

Design	*Boundaries*	*Rules*	*Monitoring*	*Rationale*
Centralized	Medium	Medium	Medium	Medium value, medium confidentiality
Decentralized	Unnecessary	Strong	Effective	High value, high confidentiality
Collective	Strong	Strong	Effective	High value, high confidentiality, small market

Although promising, these three variants are based on conceptualizations on the institutional requirements which are not fully functional. Hence, these requirements are not exhaustive as they consider only economic perspective. Apart from these, there are also additional requirements which specify further necessary aspects of data marketplace platforms.

2.4 High-Level Architecture (*HLA*) Framework

Following the decision of building the high-level architecture of a data marketplace platform, a simple framework was formulated which could help in building the same. Since our focal entity is the species of data marketplace platforms which is a technological entity, we decided the scope of the potentially resulting architecture to be technological which means that the resulting architecture would be a technological architecture of the data marketplace platforms but only representing their surface-level (high-level) information with no technical (low-level) specification. Following this scope formulation, the constituents of the framework were formulated. For any technology, the underlying principle is that the customers of the technology dictate what the technology should deliver. Hence, as a norm for developing any technology, firstly, the requirements of that technology are specified; then the profiles of the consumers who potentially use the technology are designed and finally, the surface-level (high-level) components are decided which reflect the fundamental functionalities of the focal technology satisfying its previously-specified requirements. This philosophy holds good not only for a technology but also to a wider scope till the level of organizations. Hence, this framework is not just specific to data

marketplace platform but also other business entities ranging from a simple information systems to complex organizations.

Based on the above motivation, the attributes of the framework were formalized as follows:

1. ***Functional Requirements*** specify the basic requirements which are required to ensure the basic functioning of the focal entity. These requirements can be specified at the surface level without going into any detail to support the high-level philosophy of the framework.
2. ***Customers*** signify the customers who use the offering of the focal business entity. Although a non-technological attribute, the customers form a crucial ingredient as they are the ones using the technology for their benefit. Hence, it is necessary to define the customer profiles who utilize the technology.
3. ***Functional Components*** represent the block box versions of all the components of the focal entity which embody the fundamental functionalities and features which satisfy the previously-developed functional requirements Consequently, when building the high-level architecture, the functionalities and the features have to be formulated.

Figure 2: High-Level Architecture (HLA) Framework

2.5 High-Level Architecture of a Data Marketplace Platform

The *HLA framework* was applied to build a high-level architecture of a generic data marketplace platform and the same is discussed in this section. As specified earlier, the resulting architecture will be a technological architecture of the data marketplace platform with surface-level (high-level) information with no technical (low-level) specification. The application of the *HLA framework* involves populating the values for the attributes: *functional requirements*, *actors* and *functional components* with the information either from the literature or further conceptualizations as applicable for our focal data marketplace platform (*many-to-many B2B data marketplace*).

2.5.1 Functional Requirements of the Data Marketplace Platform

Our literature survey found two sources dealing with functional requirements for data marketplaces:

- *Institutional requirements* of a data marketplace platform by Koutroumpis et al. (2017);
- *Goals* of a DMP developed by Chakrabarti et al. (2018) for an Industrial Data Space project.

These requirements were analyzed, and the appropriate ones were either adopted directly or interpreted as applicable to the focal data marketplace platforms of this research. These constitute necessary conditions for the basic functioning of a data marketplace and are listed and described as follows:

- ***Boundary Conditions***: Strict boundary conditions help in authorizing only the legitimate participants willing to share or buy data. This helps in safeguarding the data from unauthorized access from malicious sources.
- ***Data Provenance***: The lineage of data should be tracked and the change of ownership of each data point in the offering should be documented. The provenance information is the "metadata" of the data product and the platform should have a feature to manage this metadata which helps in preserving the legal usage of data.

- ***Data Governance***: This requirement is a way of governing the trading of data by having mechanisms for management and maintenance of data, traceability of data exchange and data use.
- ***Data Economy***: This requirement simply reflects the business purpose of the data marketplace platform which is to generate revenue stream for itself through its services. Usually, this is achieved through the commissions earned from the data marketplace platform services or by further additional means.
- ***Data Sovereignty***: The platform should have mechanism for the data provider to have control over his dataset, which can be enabled by handling permissions, usage restrictions, data contracts etc. or through technological solutions. By this, the provider can protect the legality of the data and not be worried about it being misused by the data consumer.
- ***Secure Data Exchange***: This is a requirement which relates to the most fundamental aspect of the data marketplace platform, the data exchange. The data exchange should happen in the most secure way because the data being exchanged is of high commercial value. The disclosure of such data will reduce its value and result in commercial, reputational and regulatory losses to the data actors. Hence, data exchange from the origin of data (data provider) to the actual point of use (data consumer) should happen in a secure way.
- ***Data Exchange Platform***: This is a complementary requirement resulting from combining all the previous requirements which is to have a fully equipped data exchange platform which could enable the data actors to trade data.

Among these requirements, the *boundary conditions*, *data provenance*, *data sovereignty* and *data governance* collectively were inspired from the institutional requirements as suggested by Koutroumpis et al. (2017); while rest of the requirements were adopted and interpreted from Chakrabarti et al. (2018).

2.6 Actors around Data Marketplace Platforms

Broadly, there can be two kinds of actors involved in using a data marketplace platform; namely, ***Data Providers*** who sell data and ***Data Consumers*** who buy the data. Owing to the scope of the focal data marketplace platforms of this research (*many-to-many B2B data marketplaces)*, the customers here comprise only of businesses who have adopted data-driven business models, but not consumers.

2.6.1 Data Providers

Data Providers are the organizations that publish and sell data on the data marketplace platform. The big data explosion has helped organizations to create business models around the data itself as an offering and reap in economic incentives (Guszcza et al, 2013). The data providers can further consist of three types of actors:

- ***Data Collectors***: They capture the data either as their main activity (e.g. meteorological measurements, web crawlers etc.) or as a byproduct from their main activities (e.g. social media, IoT services etc.). They provide raw datasets on the data marketplace platforms.
- ***Data Managers***: These are the organizations that catalogue, clean and parse the raw data into more meaningful and more-interpretable data (Leiponen et al., 2016). They basically perform data curation services like formatting, language translation, identification of outliers etc (van Bommel et al., 2005), and improve the value of the data to be traded on the data marketplace platforms.
- ***Data Aggregators***: These are the organizations that compile data from multiple sources and aggregate to create valued data products. They search, cross reference and contextualize the data to find correlations or just combine the datasets to create a differentiated data which can be useful for other businesses (Leiponen et al., 2016).

Although these customers perform different key activities, from the perspective of a data marketplace platform, they offer their data on the platform for sale. Hence, they are grouped into one data customer as *data provider.*

2.6.2 Data Consumers

Data Consumers are the organizations that search and purchase data on the data marketplace platform. Usually, these are the organizations that have adopted data-driven philosophy in their operations like in their decision making, optimizing business processes or to create data-driven products or data-driven business models (Hartmann et al., 2016). These activities fueled by the data helps the data consumers understand their customers better, differentiate their offerings to serve them better and thus, attain competitive advantage in their respective markets (Liang et al., 2018).

2.7 Functional Components of the Data Marketplace Platform

The functional components here were derived from the functional requirements derived from Koutroumpis et al. (2017), Quix et al. (2017), and the *Enterprise Data Marketplace* (EDM) by Wells (2017) in Section 2.5.1. During the formulation of these components, the underlying condition which guided the process was that all the conceptualized components should enforce all the functional requirements in a comprehensive way. Depending on the platform design of the data marketplace, the object being managed by the data marketplace platform can either be both data and metadata (centralized) or just metadata (decentralized). For simplicity's sake, and to involve both platform designs, we use the term "*(meta)data"* to represent the object being managed by the data marketplace platform when discussing the components that have common meaning for both the designs. However, when dealing with specific platform designs, the corresponding term of either *data* or *metadata* is used. The different functional components of the data marketplace platform were formulated as follows.

2.7.1 Identity Management

The *Identity Management* is responsible mainly for enforcing the *boundary conditions* for the participants to enter the data marketplace platform and access its services. A screening process can be put in place for the participants to enter the data marketplace platform in order to establish the legitimacy of that participant so that the platform services can be protected from malicious actors. After the entry, the credentials and privileges of the participants must be managed and maintained. To handle these features, three services were conceptualized as part of *identity management*; namely, ***induction, authentication*** and ***authorization***. Basically, this component takes care of the security aspects of the data marketplace platform. This component stores and manages the credentials and privileges which can be termed as the identity information of the participants. This identity information can also contain participant profiles with sensitive information like personal identifiable information, payment details et cetera which needs to be protected. Hence, *identity management* should be implemented with utmost secure technologies.

2.7.2 Broker Service

Broker Service is the most fundamental component for a data marketplace platform to have as it comprises of the features that reflect the platform aspect of the data marketplaces. *Broker Service* is responsible for two kinds of features which are described as follows:

- ***Data Management Services***: This entails the catalouging, categorizing, and tagging of data, to ensure users' have sufficient meta data to utilize the offerings (see the (meta)data inventory in Section 2.7.6.) Furthermore, the data management services include data tracking capabilities to account which data is being used, and where it comes from (data lineage.)
- ***User Interaction Services***: A data marketplace has to interact with its users (data consumers and data providers alike.) Hence, the broker service should provide some form of interface to enable this interaction. In its simplest form, this might just be a shop-like web-interface, while more complete sollutions may be integrated with other functional components, e.g., data analytics services in the form of a SaaS (Software-as-a-Service) product.

2.7.3 Backend Features: Data Management Services

The *backend features* comprise of the services which manage the (meta)data which are conceptualized as follows:

- ***Data Cataloguing***: This is a service which involves creating and maintaining the catalogue inventory of every (meta)data present on the data marketplace platform. This service basically showcases the portfolio of the data marketplace.
- ***Data Marketplace Curation***: This service involves 2 activities: *data categorization* and *data tagging*. Data can be categorized on the high level as *raw data*, *integrated data* and *aggregated data*. The data can also be categorized based on other context like, by quality, subject area, timeliness, industry etc. Data tagging complements the categories by tagging each data set helping the data consumer to find the relevant (meta)data. Overall, the categories help in arranging the data in a taxonomy helping the data consumers to browse for data while tagging enables data consumers to search for the required data. Curation further involves explicit tagging of the data that is sensitive to privacy, security, legal compliance and other constraints. The activities of data categorization and data tagging applies for both platform designs as the curation is with respect the data proposition provided by the data marketplace. For a centralized platform, the curation represents for the data that is there on the platform; whereas, for a decentralized platform, the curation represents the data being transacted over the platform by searching and selecting data based on the metadata information.
- ***Data Tracking***: This service tracks the lineage and usage of the transacted data which is appropriately updated on that data's metadata information; thus, enforcing *data provenance*.

2.7.4 Frontend Features: User Interaction Services

The *frontend features* include the services that provide a marketplace experience for the participants of the data marketplace platform. This basically include features to publish, browse, search, transform and access the (meta)data for the participants. Thereby, they enforce the fundamental platform requirement of matching data consumers to the prospective data providers to fulfil the former's data needs.
Through these features and services, the *broker service* enforces multiple functional requirements like, e.g., the *data exchange platform* specifically through the services of *(meta)data cataloguing*, *data marketplace curation and user interaction services.* Furthermore, the *broker service* enforces *data governance* through the services of backend features and *data provenance* through *data tracking* service. Overall, by providing the fundamental data marketplace platform services, *broker service* enforces the requirement of *data economy* for the data marketplace ecosystem.

2.7.5 Clearing House

Clearing House is the component essential for any digital marketplace. The component houses the repository of data exchange transactions information. Every data exchange transaction is recorded and stored in here. This component provides transaction reports essential for billing, and, help in tracing the lineage of a data product, thus enabling the requirement of *data provenance* and in turn, *data governance*.

2.7.6 Data Inventory

Data Inventory is a storage component which reflects the repository of the (meta)data. The broker service orchestrates the processes for the uploading and retrieval of (meta)data from this component. Based on the marketplace platform design, the data can either be stored at the provider site enabling the requirement of *data sovereignty* while the platform housing only the metadata inventory (decentralized platform); or both data and metadata can be housed on the platform (centralized platform). The enforcement of *data sovereignty* is weak on the centralized platform as the data providers participate only based on the intangible trust towards the data marketplace provider i.e. the data provider is expected to trust the marketplace provider's word for what has been done with the data after it left the data

provider's premises. However, the metadata maintained in the inventory also contains information about the terms of usage in a contractual form which provides control for the data provider over the usage of his/her data; thus, enforcing *data sovereignty*. Additionally, the component contributes partially towards enforcing *data governance* with the help of the *Broker Service.*

2.7.7 Data Exchange Service

Data Exchange Service comprises of the mechanism through which the physical data travels from the data provider to the data consumer in a secure way. Through the defined process, this component enforces the requirement of *secure data exchange*.

2.7.8 Data Analytics Service

Data Analysis Service is an additional way of creating value for the participants. We signify this component as the provisioning of data analytic tools which can be used to enrich the datasets into more valuable products. The tools may include data preparation, aggregation, transformation, language translation, visualization and many more. The tools can be provided on the platform in the form of downloadable software or SaaS. These tools are handy for big data players to refine their data offering and make it more attractive in the data marketplace platform. By doing so, these tools can bring in additional revenue to the data marketplace; thus, contributing towards enforcing the requirement of *data economy.*

2.8 Generic Pre-MPC Data Marketplace Platform

Using the outlined components and requirements, we constructed the ***Pre-MPC Data Marketplace Platform 1.0***, see **Figure 3**.

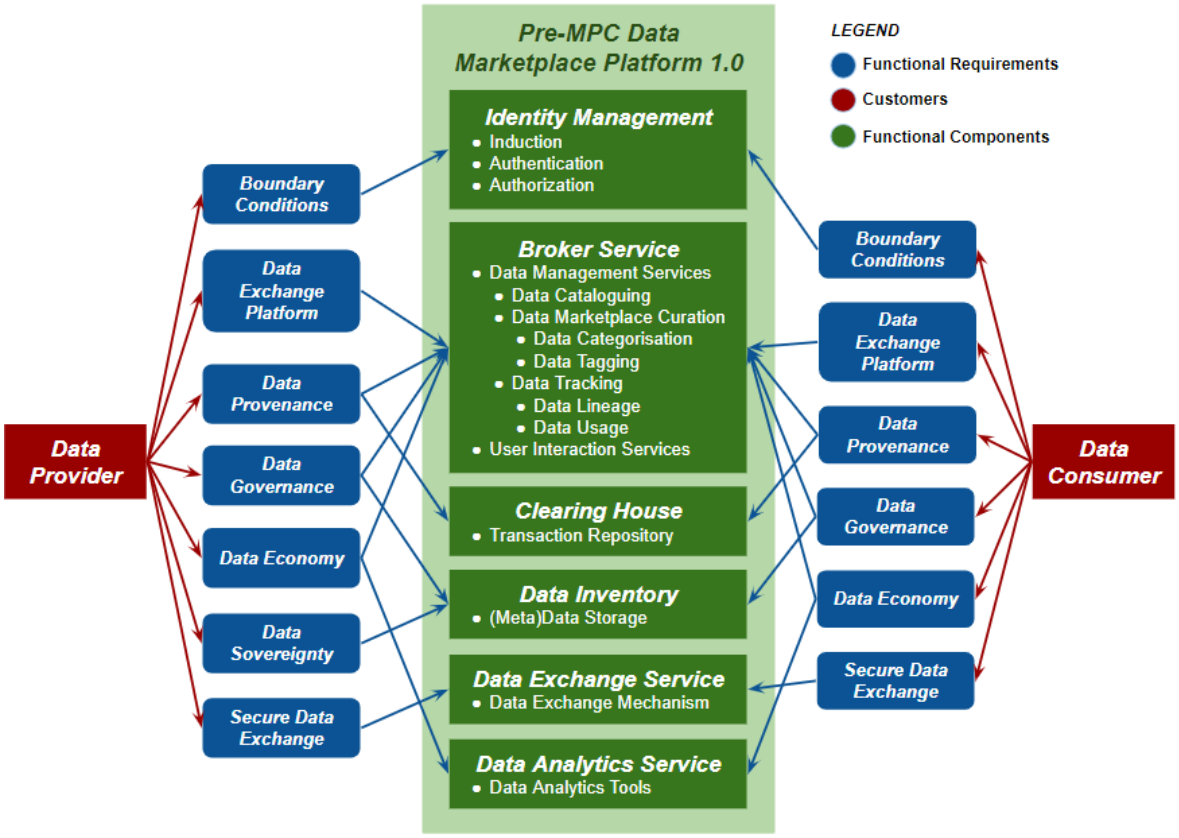

Figure 3: High-Level Architecture of a generic Data Marketplace Platform (Pre-MPC Data Marketplace Platform 1.0)

This page has been intentionally left blank.

3 Background on Threat Modelling

We first conducted a literature study on the process of threat modelling with an underlying criterion to safeguard the fundamental computer security properties: *Confidentiality*, *Integrity* and *Availability* (***CIA***). The aim of the literature study was to explore and understand the threat modelling process and further, to search for a suitable framework or methodology which might be applicable to our case. Based on existing frameworks/methodologies, a new framework was developed to appropriately perform threat modelling on the *Pre-MPC Data Marketplace Platform 1.0* from Chapter 2.

3.1 Literature Search and Selection Methodology

The aim of the literature analysis was to determine an approach to carry out threat modelling on the high-level architecture of the data marketplace platform from Chapter 2. Consequently, the focus of the literature search was for a threat modelling methodology which can accommodate a high-level architecture of a technological entity with no low-level technical specification.
With this aim, a simple search was performed on Web of Science and then Scopus with the search phrase, ***"threat modelling"*** which resulted in ***199*** and ***683*** articles respectively. The articles ranged from dealing with threat modelling of specific systems like unmanned autonomous systems; to detecting specific types of cyberattacks like Ransomware; to securing specific domains like cloud, IoT, supply chain environments etc.
Clearly, there exists a plenty of literature dealing with threat modelling of a variety of systems. Since studying and comparing each of these methodologies would evidently be a cumbersome job, the strategy was then changed to search for *review/survey* articles which dealt with the analysis and comparison of different threat modelling methodologies. This strategy was expected not only to help in finding a suitable methodology but also in covering bases of threat modelling in different areas, scopes and levels to ensure comprehensiveness of the search. Consequently, a key word search of ***("threat modelling" AND (review OR survey))*** on Web of Science and Scopus yielded ***10*** and ***53*** articles.
Out of these, ***2*** articles were identified in the results of Scopus which satisfied our focus to some extent. Firstly, "*Threat modelling – A systematic literature review*" by Xiong & Lagerström (2019) consisting a review of ***54*** articles. Secondly, "*A review of threat modelling and its hybrid approaches to software security testing*" by Omotunde & Ibrahim (2015) comprising of a review of ***101*** articles. The limitation of these articles was that both their review consisted of only software engineering approaches (technical aspects) to threat modelling. Hence, they did not fit our requirement. At this point, we broadened our boundaries of search by conducting the same keyword search, ***("threat modelling" AND (review OR survey))*** on Google Scholar in hoping to find review articles from wider range of sources.
This yielded in several results which mostly contained security requirements engineering and security practices. To include the cybersecurity aspect into the search, the key word was refined to ***("cyber threat modelling" AND (review OR survey))*** given that the threats were investigated with respect to the computer security properties (*CIA*). This resulted in a review article authored by Bodeau, Mccollum, & Fox (2018) as part of The MITRE Corporation working for the Homeland Security Systems Engineering and Development Institute (HSSEDI). The authors conducted a survey of threat modelling frameworks, analysed the methodologies, compared them, and created a framework out of the knowledge obtained from the reviewed methodologies. The article presents a comprehensive overview of threat modelling, expanding beyond software engineering approaches. As the article heavily overlaps with our own findings, it forms the basis of the following section. We amend the findings of Bodeay et al. with additional literature from our own review where necessary.

3.2 Threat Modelling Objectives

Bodeau et al. (2018) define threat modelling as "*the process of developing and applying a representation of adversarial threats (sources, scenarios and specific events) in cyberspace*". Logistically, this process can be carried out in several different ways depending on the context. Microsoft provided a fundamental

approach to serve as a starting point for the threat modelling process which was directed towards web applications (Meier et al., 2003). The steps of the process as developed by Microsoft involved: "*1) Identify security objectives; 2) Create an application overview; 3) Decompose the application; 4) Identify threats; and 5) Identify vulnerabilities*". This approach and its interpretations have been adopted and advocated by many researchers to carry out threat modelling (Steven, 2010; Kamatchi & Ambekar, 2016). EMC added an extra feature to this process in the step of identification of the threats. A library of generic threats was developed to guide the threat modelling activity which simplified the process of identification of threats in EMC's context (Dhillon, 2011). Further, the threat modelling process was adopted in the areas beyond web applications and software development. The process of threat modelling was modified according to the context of the respective areas which led to the advent of different threat modelling frameworks. Currently, the process of threat modelling involves selecting a threat modelling framework and developing a threat model by populating the framework with values as relevant to the intended context (Bodeau et al., 2018). Using the framework, we can then construct threat scenarios as representations of the identified threats to characterize ideal mitigations. Since these frameworks and their respective terminology are highly context dependent, the threat modelling process cannot be standardized. This provides a flexibility to design the threat modelling process effectively to the needs of the context and consequently, the resulting threat model would be effectively valid in that context.
It is evident that the crucial aspect of threat modelling process is the formulation of the context in which the threat modelling will be carried out.

3.3 Key Concepts and Terminology

Before diving into the aspects of context formulation, it is important to brush up on key concepts and terminology related to threat modelling. To start off right from the basics, a *model* is defined as "an abstract representation of some domain of human experience, used to structure knowledge; to provide a common language for discussing that knowledge; and to perform analyses in that domain" (Bodeau et al., 2018). The domain here is the threat landscape of cyberspace around the technological organizations. The terms used in threat modelling involve threat, threat actor, threat vector, threat scenario, attacker, attack, attack vector, malicious cyber activity, intrusion et cetera. These terms are defined differently in different threat modelling approaches based on the assumptions about the context of the technological and operational environment. However, few concepts are generally crucial to be aware of in the threat modelling area. Bodeau et al. (2018) suggests these concepts as,

- undesirable events (***threat or threat event***)
- forces or actors causing the events (***threat source***)
- structured accounts of how the event could cause the harm (***threat scenario***) and
- the resulting harm (***consequence***)

The term *threat/threat event* has different interpretations. The risk assessment guide published by National Institute of Standards and Technology (NIST) in its publication NIST SP 800-30R1 defines *threat* as *"Any circumstance or event with the potential to adversely impact organizational operations (including mission, functions, image, or reputation), organizational assets, individuals, other organizations, or the Nation through an information system via unauthorized access, destruction, disclosure, or modification of information, and/or denial of service"* (NIST, 2012). This definition provides a generic view of threat from a wider scope. A narrower definition from the perspective of the information systems literature is given by The Federal Financial Institutions Examination Council (FFIEC) Information Security Handbook on Risk Assessment (FFIEC, 2016), which reflects our focus of threat modelling, "*Threats are events that could cause harm to the confidentiality, integrity, or availability of information or information systems, through unauthorized disclosure, misuse, alteration, or destruction of information or information systems.*".
Threat sources comprise of 4 types as identified by NIST SP 800-30R1. They are: *adversarial, accidental, structural* and *environmental*. For our focal system which is a high-level abstraction, *structural sources* are irrelevant as no technical specification is available. The same goes with *environmental sources* as the focal system is a technological platform which is not *directly* affected by *environmental threats*. Although both of these sources come into picture at the further levels of threat

modelling. *Accidental* sources are the ones who intent no harm but accidentally take actions that result in harm to the system. They depent on the processes existing in the system which can accidentally go wrong. These are somewhat relevant to our context which will be explicated during the threat modelling activity. Finally, *Adversarial* sources are described as *"individuals, groups or organizations that seek to exploit the organization's dependence on cyber resources (i.e., information in electronic form, information and communications technologies, and the communications and information-handling capabilities provided by those technologies)*" (NIST, 2012). Basically, the *adversarial* sources are the ones with malicious intent that comprise of further aspects, *characteristics* and *behavior*. *Characteristics* includes further 2 aspects, *capabilities* which reflect the expertise and resources held by the adversaries and *intent* comprising of cyber goals (e.g. gaining access) or intended cyber effects (e.g. denial of service, data breach etc.); non-cyber goals (e.g. Financial gain); and risk trade-offs. *Behaviors* are described by *tactics, techniques and procedures (TTPs). "Tactics are high-level descriptions of behavior, techniques are detailed descriptions of behavior in the context of a tactic, and procedures are even lower-level, highly detailed descriptions in the context of a technique. TTPs could describe an actor's tendency to use a specific malware variant, order of operations, attack tool, delivery mechanism (e.g., phishing or watering hole attack), or exploit."* (Johnson et al., 2016). The *behaviors* of the adversarial threat agents can be characterized in terms of *threat vector* or *attack vector* they use (Bodeau et al., 2018). *Attack vectors* are *"general approaches to achieve cyber effects, and comprise of cyber, physical or kinetic, social engineering and supply chain attacks*" (Bodeau et al., 2018). *Threat scenario is* defined by NIST SP 800-30R1 as "*a set of discrete threat events, associated with a specific threat source or multiple threat sources, partially ordered in time*" (NIST, 2012). This relates to the 7 stages of hacking suggested by D. A. Smith (2017) where each stages signifies a single threat event with the whole affair translating to threat scenario.
And finally, *consequences* are the harm caused in terms of effects on information and information systems. The *cyber effects* are expressed as loss of confidentiality, integrity and availability and are translated into effects on the systems, business functions, organization and its customers.
These are some of the key concepts and terminology which are relevant to the threat modelling activity.

3.4 Threat Modelling Context

Bodeau et al. (2018) provide a summary of the context of threat modelling, consisting of three aspects, see **Figure 4Fehler! Verweisquelle konnte nicht gefunden werden.**:

- The **Scope**, i.e., *what* we look at. This ranges from the information system itself to the national or international context of a system. Classical threat modelling is often system specific, and driven by software and system engineering perspectives, as already mentioned in the high-level summary of our literature survey results.
- The **Approach**, i.e., *how* we perform threat modelling). Classical threat modelling approaches usually look at the relevant threats, the affected system, and the protected assets. Depending on the specific case, the threat model might focus more on one of these aspects. Still, commonly the focus will entail an intersection of these aspects.
- The **Purpose**, i.e., *why* we perform threat modelling. The classical purpose of threat modelling is risk management, i.e., "*risk framing, risk assessment, risk response and risk monitoring*" (NIST, 2011). Nevertheless, other aspects may be relevant, e.g., war-gaming to improve operational efficieny and general operational and design analysis.

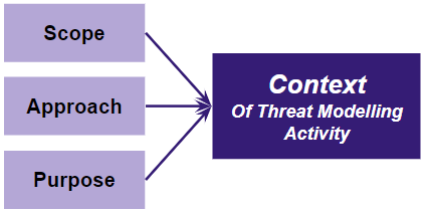

Figure 4: Conceptualization for the Context of Threat Modelling

3.5 Threat Modelling Frameworks

To gain more insight towards framing the context for our threat modelling activity, threat modelling frameworks which operate in different contexts were reviewed. Similar to Bodeau et al. (2018), the frameworks were categorized for the discussion ahead based on their purpose; i.e. for *Cyber Risk Management*, for *System Design & Analysis* and for *Threat Information Sharing*. A widely-used methodology in each category and later a few populated threat models which contain commonly identified threats already familiar are discussed.

3.5.1 Frameworks for Cyber Risk Management

There are several frameworks which help the purpose of cyber risk management. One such approach was developed by National Institute for Standards and Technology (NIST) in their various publications which contain threat modelling as an explicit part of their risk management process. As defined earlier, NIST's risk management framework contains 4 components: *risk framing, risk assessment, risk response and risk monitoring*. Thread modelling is part of their first component, risk framing. They define risk framing in their publication NIST SP 800-39 as, "*the set of assumptions, constraints, risk tolerances, and priorities/trade-offs that shape an organization's approach for managing risk*" (NIST, 2011). This step also involves assuming about the threat environment of the focal entity. The threat environment here is described as threat sources and threat events including the types of adversarial TTPs and adversarial characteristics (capabilities, intent etc). These assumptions form the threat model and the risk assessment helps in prioritising the threats and documenting them for the next step, risk response. The threat model is updated every time the risk assessment is carried out. They provide a representation threat model which comprise of; a taxonomy of threat sources with their characteristics, a set of threat events and a taxonomy of predisposing conditions which help in judging the likelihood of the threats. This initial threat model forms the starting point to start the brainstorming of the assumptions of the focal entity's context to develop its threat model. Bodeau et al. (2018) have surveyed several other frameworks and methodologies dealing with cyber risk management. Their work can be referred for more detailed analysis and relevance of the frameworks.

3.5.2 Threat Modelling for System Design and Analysis

This category contains a plenty of highly structured threat modelling approaches which supports the system design decisions and its development process. The survey article by Xiong & Lagerström (2019) as mentioned earlier consists of the analysis of 54 articles employing different methodologies which only deal with the purpose of system design and testing. Bodeau et al. (2018) also have reviewed few methodologies out of which the most popular one is reviewed here; the widely referred methodology developed by Microsoft as part of their secure Software Development Life Cycle (SDLC) agenda, ***STRIDE*** model. STRIDE is an acronym which stands for "*Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service and Elevation of Privilege*" which represent the general categories of threat vectors applicable in software environment. STRIDE primarily helps in the steps of *threat identification* of the threat modelling process proposed by Microsoft. It is flexible and highly dependent on the system specification and architecture. Each of the components and their interaction with each other and the flow of data are analyzed, and STRIDE mnemonics are applied to each component to identify threats specific to that component. Based on these findings, the developer can identify different bugs in the system and decide how to fix them. STRIDE is helpful in identifying threats in the system but further techniques like threat trees, attack trees etc. are needed to model the threat events and scenarios. The STRIDE model is like the risk framing step of the NIST framework but in a software environment. It is supported by another model called ***DREAD*** (Damage, Reliability, Exploitability, Affected Users and Discoverability) which is also developed by Microsoft to evaluate the threats and choose the relevant threats to mitigate; like risk assessment step of NIST framework. Several researchers have used and have recommended STRIDE framework to model threats in a variety of environments by customizing it to fit their requirements (Steven, 2010; Kamatchi & Ambekar, 2016; Marback, Do, He, Kondamarri, & Xu, 2013). It is important to notice that STRIDE takes *system centric*

approach to model threats for the purpose of system design and testing. A different way of system-centric threat modelling is proposed by Uzunov & Fernandez (2014) in which they decompose the system architecture into its generic functional components and develop a taxonomy of threats based on the characteristics of each component. The taxonomy is used as a reference when the threat assessment is carried out for specific systems and the newly identified threats are updated in the taxonomy. This is the most applicable way of doing system-centric threat modelling, but it requires complete specification of the system and an expert threat modeler. There are many more methodologies which take different approaches of threat modelling to system design. For example, Intel's Threat Agent Risk Assessment (TARA) which takes the threat-centric approach (Rosenquist, 2009) and IDDIL/ATC methodology which takes an asset-centric approach with the first step being to identify and characterize the assets in the context (Muckin & Fitch, 2017).

3.5.3 Threat Models for Threat Information Sharing

The threat modelling frameworks discussed so far can direct the process towards developing the threat models. They were used in the initial years when threat modelling was an infant field to research. But since then, the field has evolved, and more sophisticated techniques have been developed to carry out threat modelling. As a result, the previously discussed frameworks are often not used in the organizations. From the representation threat model of NIST SP 800-30R1, organizations develop hybrid or customized approaches for various purposes suited to their business processes. Here, some of the threat models are discussed which were developed for various purposes but help by lending the information about a variety of techniques used by threat actors in different environments. Bodeau et al. (2018) identified a few threat models which include 2 kinds: enterprise-neutral and enterprise-oriented threat models.

3.5.3.1 Enterprise-Neutral Threat Models

Enterprise-neutral threat models consist of adversary characteristics and behaviors consisting of attack techniques within a general technological environment. The focus here is only on the threat event with adversary techniques and do not incorporate information about enterprise characteristics like its architecture, assets and systems. Basically, they take a threat-centric approach. Some of the examples include ATT&CK (Adversarial Tactics, Techniques & Common Knowledge), CAPEC (Common Attack Pattern Enumeration and Classification), OWASP (Open Web Application Security Project) etc. ATT&CK is developed by the MITRE Corporation (The MITRE Corporation, 2015) and provides an account of adversary behavior within an enterprise network i.e. post-access through a successful entry exploit (Bodeau et al., 2018). ATT&CK consists of a repository of adversary attack techniques which operate in a network powered by Microsoft Windows environment. The repository consists of 10 categories of tactics with each tactic containing a list of attack techniques and potential mitigations. The tactic categories are: persistence, privilege escalation, defense evasion, credential access, discovery, lateral movement, execution, collection, exfiltration and command & control. Like ATT&CK, CAPEC model provides a catalogue of attack patterns with more detail than ATT&CK which help in categorizing the attacks in a meaningful way; OWASP comprises of 12 categories of attacks applicable in web applications. These models lend several categories of adversary TTPs which can be used to model the threats in the realm of the focal context.

3.5.3.2 Enterprise-Oriented Threat Models

These are the threat models generated after the threat assessment of particular enterprises. Since the models contain sensitive enterprise-specific information about the ways it could be attacked, these are generally not shared. However, Bodeau et al. (2018) identify 3 generic models in this category which indirectly deal with enterprise-specific threat modelling. One of these models is MITRE's Threat Assessment and Remediation Analysis (TARA) which will be discussed here in brief.
MITRE's TARA is actually a methodology developed for identifying threats to a system and determine countermeasures (Wynn, 2014). The threat identifying component of the MITRE's TARA is called Cyber Threat Susceptibility Analysis (CTSA) which identifies and evaluates potential cyberattack events and patterns. Like the previously discussed populated threat models, CTSA also builds a threat

catalogue focusing on the attack vectors. Additionally, MITRE's TARA contains a taxonomy of vector groups and a set of tools which map the attack vectors to different system environments and technologies. This is the differentiating feature of MITRE's TARA compared to other methodologies. MITRE's TARA proposes a threat modelling process like that of NIST SP 800-30 and Microsoft: (i) identify the scope, architecture and technological components; (ii) make assumptions about the types of adversaries and techniques; (iii) identify the threats appropriate to the scope and assumptions. MITRE's TARA also has its own way of assessing and prioritizing threats to mitigate. To sum up, MITRE's TARA, it is useful for threat information sharing as it contains a catalogue of attack vectors and tools to map them to the system environments.

Table 4: Reflections on Different Threat Modelling Frameworks

Framework	*Scope*	*Approach*	*Purpose*	*Technical Expert needed?*
NIST SP 800 30R1	Organization, Mission, System	Flexible – Can be made Threat, System or Asset centric based on the information available	Risk Management	Technical Expert translates to more Validity
STRIDE	System	System-Centric	System Design Analysis	Depends on the context specification
Uzunov & Fernandez (2014)	System (Distributed networks only)	System-Centric	System Design Analysis	Yes
Intel's TARA	Organization, Mission, System	Threat-Centric	System Design Analysis	No
IDDIL/ATC	System	Asset-Centric	Risk Management	No
ATT&CK	System (post network entry)	Threat-Centric	Threat Information sharing	Depends on the context specification
CAPEC	System	Threat Centric	Threat Information Sharing	Depends on the context specification
OWASP	System (Web applications only)	Threat-Centric	Threat Information Sharing	Depends on the context specification
MITRE's TARA	Organization, Mission, System	Partly System-centric and partly Threat-centric	Risk Management and Threat Information Sharing	Yes, but a lesser expert compared to other expert methods

3.5.4 Reflection on the Frameworks

Firstly, the risk management framework developed by NIST in their publications are very generic. The method is highly flexible and depends on the threat modeler to define the detailed tasks as the framework just motivates the threat modeler with relevant aspects; but the methodology does not direct him/her with detailed tasks. Hence, this framework can act as a starting point to learn different aspects of threat modelling, but the threat modeler should be aware of the detailed information of the scope in which he is operating to form concrete assumptions to start the threat modelling process. However, the catalogue of different taxonomies helps in the step of threat identification. Since the methodology is generic and involves lot of assuming and conceptualizing, the threat modelling can be time consuming and a tedious process; and needs managers and technicians working together. Then, we discussed STRIDE. It takes a software engineering and system centric approach. Evidently, it can only be carried out by an expert technician and the manager has a lowest role to play in the activity. Although its categories are high-level, it provides a starting direction to decide on security aspects of system design. The reflection here is like that of NIST with one exception that STRIDE applies only in the scope of information systems. On the other hand, the threat modelling methodology of Uzunov & Fernandez (2014) is specific and provides concrete steps to carry out the threat modelling which makes the methodology straight forward; but the framework applies specifically to the distributed networks and also it needs a technically expert threat modeler. **Table 4** lists the above-discussed threat modelling frameworks and threat models with the characteristics of their respective contexts.

From the above reflection, it is understood that the level of detail in which the context is described, dictates the specificity of the threat modelling process. For a context described in great detail, a specific threat modelling process can be tailored which would be highly-structured and highly effective within the context because of which relatively less expertise is needed as the validity is guaranteed by the process itself. Furthermore, the duration of the process depends on the context. Some of the examples here include: Uzunov & Fernandez (2014), Socio-technical Framework by Sabbagh & Kowalski (2015) etc. On the other hand, if the context is not available in detail, then the validity and the time duration

depend on the expertise of the threat modeler who needs to make informed assumptions about the context to carry out the threat modelling with generic methodologies. The examples here are: NIST, STRIDE, MITRE's TARA, IDDIL/ATC etc.
Ultimately, we can deduce that the process of threat modelling is highly dependent on the context and to what extent of detail it is available and described. The more detailed the description of the context, the more effective the threat modelling process becomes and the more valid the developed threat model becomes.

3.6 Context of our Threat Modelling Activity

Based on the understanding of the different threat modelling frameworks and threat models, the context of the threat modelling for the *Pre-MPC Data Marketplace Platform 1.0* from the Background on Data Marketplaces was formulated and is presented here by specifying it in the language of threat modelling: scope, approach and purpose.
Firstly, the ***scope*** was established. Since, the focal entity, the data marketplace platform is a technological platform represented with a high-level architecture, the corresponding functional components can be considered as individual information systems which can be implemented with technology alone without any human actor needed. However, there is not technical specification of these information systems but on the contrary, only features are specified which translate to the business functions of those information systems. Hence, the scope of the threat modelling activity was formalized to be at the level of ***business functions***.
Furthermore, the kind of threats and the detail in which the threats are described should be established. As mentioned earlier a threat can be described in three levels of detail namely, tactics (high-level), techniques (medium) and procedures (low-level); all of which are represented through attack vectors which can be described appropriately at three levels. Since, the unit of analysis here is only the technological components with just business functions and no technical specification, the threats were decided to be described with attack vectors (cyberattacks) at a high-level, which reflect the adversarial dimension of threat source. Related to the other threat sources, they are included when they apply during the threat modelling process. The principle behind the kind of threats is to find the cyberattack vectors applicable which are described later at high-level (*tactics*). For example, a distributed Denial of Serivce (dDoS) attack on a Server signifies a *high-level* description of threat while the whole logistics of that dDoS attack used on a specific server which entails every step involved in the attack process reflects a *low-level* description of the threats. This implies that the threat modelling activity could be performed by managerial level expert with no need for technical experts.
The ***approach***, as we have explained it, should comprise of the information we know about the focal entity, and the rest of the aspects are to be assumed in the threat modelling process. On these lines, the absence of the technical specification eliminates the system-oriented approach while the already decided specification of the tactic-level (high-level) handling of threats eliminates the threat-oriented approach. The information we do know about the data marketplace platform is with respect to the functional components and their business functions. Consequently, the assets associated with those business functions can be modelled first, to determne the whole threat modelling activity, including making assumptions on systems and threats. Hence, our approach is ***asset-centric***.
Choosing the ***purpose*** was a straight forward, as the aim is to identify the threats associated with data marketplace platforms, and to understand the threat landscape of data marketplace platforms. Hence, our purpose is the ***risk framing*** step of *risk management.*

3.6.1 Implication of the Context Formulation

Essentially, the context can be represented by a single statement as, ***"to establish the assets associated with the business functions of each functional component of the high-level architecture of a technological entity and later, assume a system specification on which applicable cyberattack vectors (described at a high-level) can be identified"***. As seen in **Table 4**, there was only one framework which satisfies our context, *NIST SP 800 30R1*. However, this framework entails the necessity of a technical expert who can make credible assumptions about the assets in the functional components, such that a

valid threat model can be generated. We do not possess this expertise as we are not technical individuals. The other option is to build a valid threat model is to have a *specific* framework applicable to the threat modelling activity for our context. There is no such framework as most of the threat modelling literature is directed either towards the software engineering area (technical) or the ones whose context are specified in a great detail to the level of infrastructure and practices in an organization. As a result, there exists a gap in the literature related to the threat modelling at the scope of business functions for the technological entities.

3.7 NGCI Apex Classification of Cyber Threat Models

Bodeau et al. (2018) addresses the problem that there is no threat modelling framework or methodology which could comply to all the contexts. He further stresses on the need for a threat modelling framework which can be customized to different purposes and used at multiple levels and scales. For this agenda, Bodeau et al. (2018) conceptualized a classification containing threat models in which the threats are described at all the levels in respective threat models (*tactics, techniques and procedures*). The classification was done as part of their NGCI Apex Program and it contains three threat models: ***High-Level Threat Models, Detailed Threat Models*** and ***Instantiated Threat Models.*** They are described as follows and the kind of threats dealt in each threat model is illustrated in **Figure 5**.

- ***High-level Threat Models:*** These contain threat events described in general terms which support high-level or sector wide risk assessment, cyber war-games or technology profiling and foraging.
- ***Detailed Threat Models:*** These support technology evaluation in which threat events are described with a little more detail in terms of specific systems, technologies or targets.
- ***Instantiated Threat Models:*** These are low-level threat models containing detailed threat scenarios which help in developing detailed cyber playbooks. These models are dependent on the system architecture and hence, these models are usually developed by the organizations themselves and are not shared to the external entities like academia since they contain sensitive information.

The block of ***Threats of Concern*** in **Figure 5** suggests generic threat events, brief narrative threat scenarios and adversary characteristics which are driven by assets. Hence, the conceptualization of the *High-Level Threat Model, comprising of cyberattack vectors described on a high-level which affects the assets,* was considered as the reference to build our threat model as it relates to our context.

3.8 HLTM Framework

We decided to design a framework specifically applicable to our context which is, *to establish the assets associated with the business functions of each functional component of the high-level architecture of a technological entity and later, assume a system specification on which applicable cyberattack vectors (described at a high-level) can be identified*. We rephrase this context into "***performing high-level threat modelling of the high-level architectures of the technological entity***" because it is driven by the concept of *high-level threat models* from the NGCI Apex Classification and the focal system, which is the high-level architecture of any technological entity. Consequently, the framework was named as ***High-Level Threat Modelling (HLTM) Framework.***

The philosophy of the framework is to break down the focal high-level architecture of the focal technological entity into its functional components, identify the business functions associated with the components and identify the threats which affect those identified business functions. The framework gives a simple structure to identifying the *high-level threats of concern* to the technological entities. Essentially, to represent in the language of threat modelling, the framework operates in the context of asset-centric (*approach*) threat modelling of the business functions (*scope*) to deduce the risk (*purpose*) associated with the focal entity. The framework consists of 6 constructs: ***Functional Component, Business Function, Threat, Cyber Effect, Business Consequence*** and ***Mitigation Technique.***

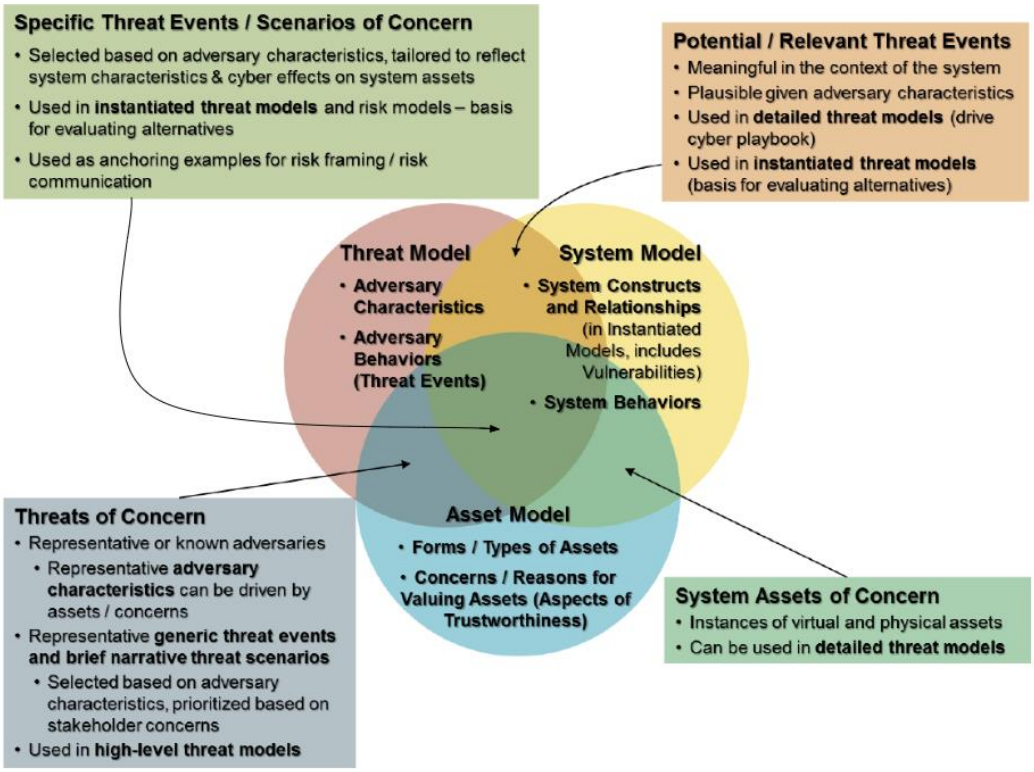

Figure 5: Types of threats in the Threat Models of NGCI Apex Program
Source: Bodeau et al. (2018)

3.8.1 Functional Component and Business Function

Both the constructs, functional component and business function, constitute the *asset* dimension of the framework. An asset is an entity which is a constituent of the system responsible for its value. In the information systems environment, asset can be defined as "*any data, device or other component that supports information-related activities, which can be illicitly accessed, used, disclosed, altered, destroyed and/or stolen result in loss*" (Jones, 2005). The choice of the assets for consideration in the threat modelling process is dictated by the business functions associated with each functional component in the high-level architecture.

Firstly, the functional components are mapped to their respective business functions. A functional component can be responsible for multiple business functions. Then, for each business function, a basic system asset is assumed which fundamentally enables the respective business function. This step signifies assuming of the *system* dimension of the threat modelling activity. The mapping gives a baseline of the low-level technical specification for which threats could be identified.

In our case of technological platform, the assets can be attributed to IT systems. The IT system assets can have different characteristics. The Single European Sky ATM Research (SESAR) differentiates 2 categories of IT assets in their Security Risk Assessment Methodology: Primary and Supporting assets (Marotta et al, 2013). They characterize primary assets as the intangible functions, information, processes, services and activities. Supporting assets are the tangible systems or components which contain vulnerabilities through which a threat agent can attack and compromise the primary asset; for example, websites, communication channels, database, server etc. We incorporated this concept of IT assets (*primary asset* and *supporting asset*) to consider assets associated with business functions.

3.8.2 Threats

By threats in this framework, we refer to cyberattack vectors which are just mentioned at high-level instead of describing the logistical process of the cyberattack happening to the assumed IT system asset.

The cyber threats generally revolve around the vulnerabilities in the system and the cyberattacks which take advantage of the said vulnerabilities. Since the system under evaluation is a high-level architecture with no technical specification, the vulnerabilities are excluded from the framework. Instead, each IT asset is considered for each business function and the cyberattack vectors appropriate to the focal IT asset is identified and attributed as its threat.
A cyberattack on a broader perspective, generally consists of seven steps which are listed in **Table 5** (D. A. Smith, 2017). Each of these steps can involve intermediate attacks which form the building blocks to a broader cyberattack. These cyberattacks are listed in **Table 6** which form the representative values that can be used during threat modelling. The list is not exhaustive and other cyberattacks can also be included for threat modelling appropriately.

Table 5: Seven steps of a Cyber Attack
Source: D. A. Smith, (2017)

Steps	*Description*
Reconnaissance	Before a full-fledged cyberattack, the attacker identifies a target and explores the information related to the target.
Scanning	After the identification of the target, the attacker searches for vulnerabilities by scanning the systems through attacks like resource enumeration and browsing.
Access and Escalation	Once the weak spot is identified, then attacker tries to gain access to the system and then escalate the privileges to move freely with the system environment. Ex: Password attacks
Exfiltration	The attacker now attempts to access sensitive assets like data and tries to extract it. Ex: Storage attacks
Sustainment	The attacker seeks to remain undetected and have unrestricted access by installing malicious programs like root kits which allows the attacker to return as and when desired.
Assault	Now, the attacker can sabotage the system either by modifying the system or disrupt it entirely by disabling it. This means the attacker has full control of the system and it is too late to defend it.
Obfuscation	This step happens when the attacker leaves a signature behind in the system to brag about his/her conquests. This usually involves confusing or diverting forensic investigation through log cleaners, spoofing, misinformation, zombie accounts, Trojan commands etc.

Table 6: General Cyber Threats to IT systems

Cyber Threat	*Description*
Botnet	A botnet is a network of remotely controlled machines used to launch wide-scale denial of service attacks against specifically targeted resources (Zhang et al., 2011).
Denial of Service (DoS, DDoS)	A Denial of Service attack consists an attempt to impeach users from accessing data or services provided by an information system (Zlomislic et al, 2014).
Eavesdropping/ Traffic Analysis	This is a form of attack where the attacker attempts to capture and analyze network data packets in the communication channel in order to identify any information that may be relevant for other types of attacks.(Fu, 2005)
Injection attacks	This attack refers to a broad class of attack vectors through which the attacker injects malicious input to a program. Particularly, SQL injection attack is considered very dangerous as the attacker can gain access to the database with sensitive data by injecting malicious value at the input field (Muscat, 2019).
Malicious code/ Payload	This is a generic family of attacks all of which involve harmful code or script designed to be executed by programs, operating systems, web servers, and any other IT device, resulting in undesired effects. These are usually carried by viruses or worms (Al-Mohannadi et al., 2016)
Man-in-the-Middle	This form of attack is a specific case in the eavesdropping type of attacks, in which the attacker interposes between the sender and the receiver and misleading them into believing their communication line is direct and secure This allows to either intercept confidential information or altering it unknowingly to the legitimate communication participants. This attack affects the confidentiality and integrity of the data in the communication channel (Conti et al., 2016).
Password attacks (Brute-force, Dictionary, Cookie Replay)	In this form of attack, the attacker attempts to identify a password or an encryption key through exhaustive checks or through cookie information from the browser until the correct string is identified (Hansman & Hunt, 2005).
Resource enumeration and browsing	This is a type of attack through which the threat actor is able to obtain from a targeted system the list of the resources that are present in the system, therefore enabling the threat actor to refine the targeting process of such resources and their consequent browsing (OWASP, 2018).
Malware/Viruses	Viruses and malware are types of malicious code/payload with various objectives, among which can be mentioned replication, data manipulation or destruction etc. (Bishop, 1991)

3.8.2.1 CIA Violations

Information security objectives are represented on a high level with the triad of computer security properties – CIA*: Confidentiality*, *Integrity* and *Availability*. We use the same triad to represent the computer security objective violated for the focal IT asset by each cyber threat identified in the previous step. The properties are described as follows,

- ***Confidentiality***: The property that the information and the services should be made available to only authorized individuals, entities or processes.

- ***Integrity***: The property of safeguarding the accuracy and completeness of information assets.
- ***Availability***: The property of information assets to be accessible and usable upon demanded by an authorized entity.

Each threat to the IT system assets results in a degradation to one or more of these properties.

3.8.2.2 Business Consequence

Business Consequence construct describes the adverse effect caused by the threat to the focal business function under consideration or to the whole technological entity in general. This construct helps in representing the adverse effects of the threats in the language of the business aspects, as opposed to the computer security properties (CIA) mentioned in the previous step. The business consequence construct can have variety of values ranging from financial loss, reputational loss, functional loss, regulatory impacts or environmental loss. Although, the value to be filled here is highly dependent on the business function under consideration.

3.8.3 Mitigation Techniques

This construct completes the circle of the whole threat modelling activity by recommending the appropriate security techniques which can mitigate the identified threats. The mitigation techniques can comprise of concrete mitigation technologies, protocols, policies and security procedures. The common security controls used are listed on a high-level by (Northcutt, 2018) in his white paper published as part of research at SANS institute. These are: *Security Awareness Training, Firewall, Anti-Virus, Intrusion Prevention System, System Monitoring, Intrusion Detection System and Encryption*. In addition to these, any other techniques and to any extent of detail can also be used to populate this construct.

3.8.4 Reflection on the HLTM Framework

The framework satisfies threat modelling context by the constructs, *Functional Component* and *Business Function* constitute the ***asset*** dimension; the assumption of basic IT system asset specification reflects the ***system*** dimension; and finally, *Threats, Cyber Effect & Business Consequence* constitute the *threat* dimension.

Furthermore, it can be deduced that all the threats and their respective business consequences to each business function reflects the high-level overview of the threat landscape around the focal technological entity. We termed this conceptualization as ***High-Level Threat Landscape*** of the focal technological entity owing to the *high-level philosophy* dealt so far. The conceptualization and the resulting HLTM Framework are illustrated in **Figure 6** and **Figure 7** respectively.

Figure 6: Conceptualization for the Threat Landscape

Because of this conceptualization, the framework was deemed fit to be applied on the high-level architecture of the data marketplace platform from Chapter 2 as it could result in an overview of the threat landscape of the *Pre-MPC Data Marketplace Platform 1.0.*

However, because of the mapping of business functions to the basic IT assets i.e. the baseline low-level technical specification, the resulting threats and their respective business consequences represent only the *baseline threat landscape of the focal technological entity*. This can be attributed as a limitation of the HLTM framework. This situation can be improved with multiple iterations of threat modelling when more knowledge is learnt on the low-level technical specification of the focal technological entity, which would further result in the *low-level threat landscape*.

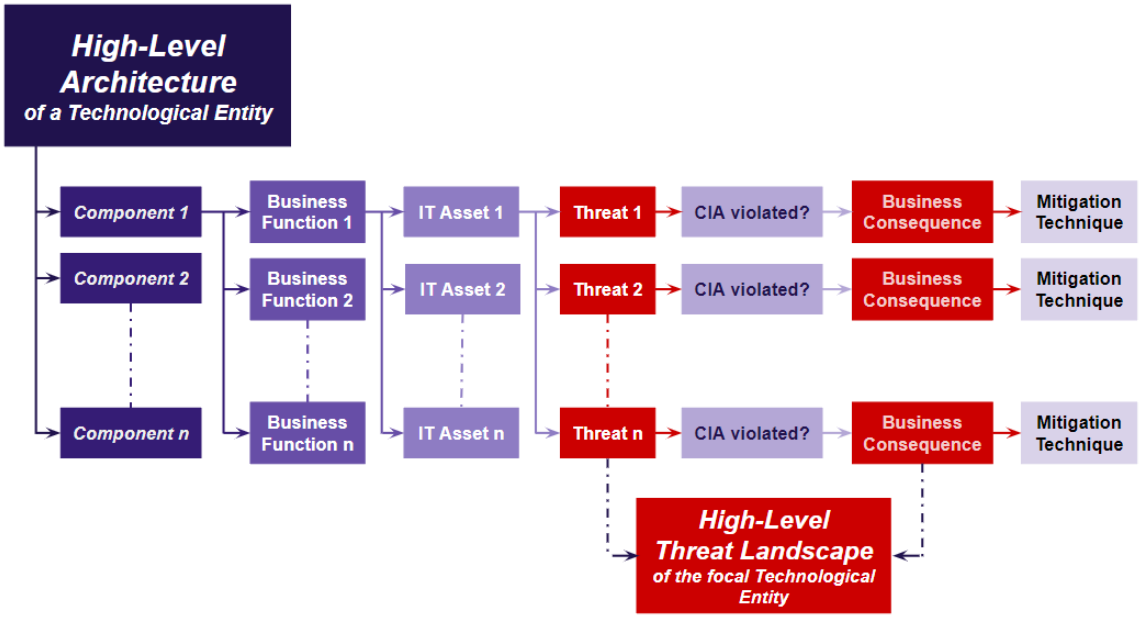

Figure 7: High-level Threat Modelling (HLTM) Framework

Apart from our research objective, the *HLTM framework* is a valuable addition to the family of threat modelling frameworks as there is none existing to address the context it is operating in, which is *to perform high-level threat modelling for the high-level architectures of the technological entities*. All the constructs in the framework are operating according to the *high-level philosophy* with almost no technical specification of the system required. Hence, the identification of the threats can be done even by a manager. However, the framework demands some basic level of technical expertise of cybersecurity which comes handy during the application of the framework like, assuming the supporting assets in IT asset stage, knowledge of which cyber threats could affect what kind of systems et cetera. Essentially, the greater the technical expertise of the threat modeller, the higher is the validity of the threat model. In that case, threat modelling by a technical expert results in a more valid threat model. In addition to this, the framework can be used to perform low-level threat modelling of specific technical architecture of technological entities. The only change would be to instead of assuming the IT assets the constituents of the available low-level architecture can be filled in the IT asset construct. Hence, the framework is flexible enough to adopt between high-level and low-level threat modelling.

4 Threat Modelling for Data Marketplaces

In Chapter 3, we developed the HLTM framework to assess IT security threats in high-level architectures. We will now apply the HLTM to the high level data market place we developed in Chapter 2 to identify the threats associated with a data marketplace platform.

4.1 High-Level Threat Model for the Data Marketplace Platform

In the larger research gap of the realization of the data marketplaces, one of the gaps is with respect to the threats faced by them. Researchers have discussed the legal and economic challenges of setting up a data marketplace (Koutroumpis et al., 2017). But the threat landscape of the data marketplaces has never been explored although it represents a significant element (security aspect) towards their realization. The researchers have touched upon such security aspect by just suggesting that the confidentiality and privacy of the data being transacted are the concerns to be explored. We went beyond this and built a comprehensive threat model comprising of all sorts of threats applicable to the high-level architecture of the data marketplace platform, thus providing a high-level overview of the threat landscape of the data marketplace platforms.
The application of the HLTM framework is straightforward as discussed in Chapter 3. Firstly, each functional component was mapped to its business functions. Then, the basic IT system assets which enable the business functions are assumed based on our experience with computer science and engineering background. IT system assets are assumed according to the template of the primary and supporting assets. The cyberattack vectors which could affect the identified IT assets are identified based on literature analysis and web search. Subsequently, the computer security property (CIA) violated and then, the consequence of the cyberattack to the focal business function or the whole entity are deduced. Finally, the appropriate mitigation technique is proposed for each cyberattack based on the literature analysis and web search. The resulting threat model represents the *Pre-MPC Threat Model 1.0* and the values of the *threat* and *business consequence* reflects the high-level overview of the threat landscape of the data marketplace platforms.

4.1.1 Identity Management

As discussed in Chapter 2, the main objective of the component, *Identity Management*, is to enforce the *boundary conditions*. This involves establishing strict processes to incentivize the customers to use the platform services and restrict access to unauthorized entities. Consequently, we established that this component involves the following business functions: *induction, authentication and authorization*. Each business function is dealt, and the corresponding threats and business consequence are discussed in the rest of the subsection.
Induction function is responsible for carrying out the screening process of the potential customers. The goal of this business function is to allow only legitimate customers to sign up for the services of the data marketplace platform. Since it is a B2B entity, the screening process should focus on establishing the legitimacy of the organization willing to sign up. The basic specification of *induction* could be that the customer must fill in the profile information on a web form which is submitted on the website. Further, the organization legitimacy could be validated by verifying its legal status with the national commercial registry database. After the verification, the customer can access the platform. The primary assets here could be the customer organization's profile information and legitimacy verification service. The supporting assets enabling the business process are the web form, the communication channel and the identity database. The identified threats are listed in ***Table 7***.
Coming to the authentication function, the assets relevant here are customer credentials and the authentication service. These could be supported by the website of the data marketplace. The threats relevant in this area are password attacks and denial of service attacks. These threats could be overcome respectively by imposing a strong password policy, and system monitoring to differentiate illegitimate requests, say from botnet, followed by tagging and isolating the source of illegitimate requests. These are listed in **Table 8**.

Table 7: Threats: Induction of Customers

Primary Asset	*Supporting Asset*	*Threat*	*CIA violated?*	*Business Consequence*	*Mitigation Technique*
Customer Organization's Profile Information	Web Form on the website	• Identity Spoofing • Masquerading	Confidentiality of the DMP services	Induction of malicious entities as customers	2-step verification of authenticity
	Identity Database	• Database Injection Attack; • Malware	CIA of the customer identity information	Compromise of authentication service through disclosure of credentials and the services of the DMP to the attacker	• Usage of secure stored procedures over direct querying; • Anti-Malware
Customer Validation Service	Communication Channel	Eavesdropping/ Traffic Analysis	Confidentiality of profile information	Disclosure of the sensitive customer profile information	Encryption
	Verification of the website of the customer organization	Counterfeit website by attacker pretending to be a customer	Integrity of the verification service	Induction of malicious entity as the customer	Verification of certificates of the consumer organization website

Table 8: Threats: Authentication

Primary Asset	*Supporting Asset*	*Threat*	*CIA violated?*	*Business Consequence*	*Mitigation Technique*
Customer Credentials & Authentication service	Website	Password Attacks • Brute Force Attack • Dictionary Attack • Cookie Replay Attack	Confidentiality of the DMP services	Access of the DMP services to malicious entities	• Strong Password Policy • Cookie Management
		Denial of Service Attack (DoS, DDoS, Botnet)	Availability of the DMP	Inability for legitimate customers to access DMP	System Monitoring for illegitimate requests

Authorization involves providing appropriate privileges to the applicable customers. This includes differentiating the customers and enforcing boundaries between the customers who have access to the platform services and the ones who have access to the data products that they have bought. The data products that are bought could be a one-time supply or a periodic supply or a real time continuous one. Depending on these parameters, the privileges should be managed and maintained. Configuration errors here might result in access to unauthorized entities. This can be overcome by periodic review of privileges and access controls. These access controls and privileges could also be target for external attackers to gain access to the system. These could be combatted with firewall and intrusion prevention system. This discussion is listed in ***Table 9***.

Table 9: Threats: Authorization

Primary Asset	*Supporting Asset*	*Threat*	*CIA violated?*	*Business Consequence*	*Mitigation Technique*
Customer Privileges	Authorization systems	Configuration errors caused by human errors	Confidentiality of the DMP services to unauthorized entity	Access of the DMP services to malicious entities	Periodic review of access controls and privileges
		Manipulation of privileges by attacker after entering the system	Integrity of authorization system	Privilege allocation and access controls to malicious attacker	• Firewall • Intrusion prevention system

4.1.2 Broker Service

The broker service component aims to provide the platform services to the customers through its 2 business functions: *Data Management* and *User Interaction*.
Data management service takes care of the background processes responsible for providing the data marketplace platform services: Data Cataloguing, Data Marketplace Curation and Data Tracking. These services could be carried out on a server which is supposed to be up and running 24/7. Threats here are that the integrity and availability of the services is disrupted, sabotaging the broker operations. One of the attack vectors capable of causing this is malware. Malware attacks comprising of Viruses, worms, payloads with malicious code can manifest into processes which could disrupt the backend services potentially sabotaging the platform. This could be combatted with an updated anti-malware installed in the system along with firewall and intrusion prevention system. In addition to this, resource enumeration & browsing attack could cause damage to data management activities by disclosing the inner mechanism of the data management services to the attacker. With this attack, the attacker can learn about the resources and their configuration to plan a follow-up sophisticated attack to the systems. This could be overcome by installing a firewall with intrusion prevention system to monitor and restrict the unauthorized requests to the system. The above discussion is listed in **Table 10**.

Table 10: Threats: Backend features: Data Management

Primary Asset	*Supporting Asset*	*Threat*	*CIA violated?*	*Business Consequence*	*Mitigation Technique*
Data Cataloguing, Data Curation, Data Tracking Services	Server in a data center with the applications carrying out data management services	Malware attacks to sabotage the DMP service	CIA of the platform services	Failure of platform services	Anti-Malware with updated malware definitions
		Resource enumeration & Browsing attack	Confidentiality of Backend resources and operations	Disclosure of the backend resources to the attacker	• Firewall • Intrusion prevention system

Frontend features involve the interface services for the customers which provide them with the data marketplace experience. All the services could generally be provided through a website and the services include publish, browse, search, transform and access the (meta)data. The threats here generally could involve the ones that affect the web applications. Open Web Application Security Project (OWASP) have researched extensively on the threat events to web applications and have published 20 threat events directed towards a number of specific web application vulnerabilities (Watson & Zaw, 2018).
All the threat events mentioned in OWASP application apply here as it is web-based service but again, the threats are implementation dependent. We included a few general threats we think are crucial. Alteration attack involves tampering the source code of the website and affect its integrity to either disrupt the service or to launch a further attack. These could be restricted by safeguarding the source code from modification which links to privilege management. Further, the usual culprits affecting the CIA apply here.
Denial of Service using Botnet attack vector could affect the availability of the website and frontend services to the customers. Eavesdropping/Traffic analysis and Man-in-the-Middle attacks could be used to intercept the information being transmitted in the communication channel from the website to the server or vice versa. Furthermore, the intercepting entity could alter the information to make malicious requests posing as a legitimate entity potentially disclosing sensitive information or sabotaging the Data Marketplace services. These could be overcome by encryption of the communication channel and valid certification of the website to establish the trustworthiness of the website. These threats are listed in **Table 11**.

4.1.3 Clearing House

The IT asset involved in this component is transaction management service which stores all the information of all the transactions happening on the data marketplace platform. This could basically be powered by a database management system and hence, the threats that apply here are database threats. These are listed in according to their applicability with the transaction management in **Table 12**.

Table 11: Threats: Frontend features: User Interaction

Primary Asset	*Supporting Asset*	*Threat*	*CIA violated?*	*Business Consequence*	*Mitigation Technique*
User Interaction Services	Website with outward facing services	Website defacement attack with alteration/ modification attacks	Integrity of the website	Faulty website with faulty functionalities resulting in reputation loss.	Restricted access to the website source code
		Denial of Service attack (DoS, DDoS, Botnet)	Availability of the website to the customers	Disruption of the website service to the customers	System Monitoring for illegitimate requests
	Communication Channel	Eavesdropping/ Traffic Analysis	Confidentiality of transmitted information	Disclosure of sensitive information	Encryption
		Man-in-the-Middle Attack	Integrity of the information and the service	• Manipulation of the sensitive information • Disclosure of sensitive information to malicious attackers posing as legitimate customers	• Encryption • Firewall • Intrusion Prevention System

The compromise of transaction management service could impact the data marketplace operations to a great extent as transaction management is responsible for the core business of the data marketplace. A compromise might lead to loss of transaction information potentially losing the track of data product being transacted. This could potentially make the platform lose the legal tracking of the product thus leading to regulatory complications. The transaction management could be safeguarded with anti-malware, firewall and intrusion prevention system to prevent external attacks while carrying out periodic maintenance and database auditing to monitor its functioning.

Table 12: Threats: Clearing House

Primary Asset	*Supporting Asset*	*Threat*	*CIA violated?*	*Business Consequence*	*Mitigation Technique*
Transaction Management Service	Database Management	Injection Attack	CI of the transaction data	• Loss of data provenance • disclosure of customer profile information with transaction details	Usage of secure stored procedures over direct querying
		Malware	CIA of the transaction data	Disruption of the website service to the customers	Anti-Malware
		Update Errors, Incomplete transactions	Integrity of the transaction data	Loss of *data provenance* losing legal connection with the data product	Frequent Auditing of database processes

4.1.4 Data Inventory

Data inventory is the storage component of the data marketplace platform which manages and maintains the data products being transacted on the platform. Based on the design of the marketplace (*centralized and decentralized*), the data inventory differs with its implementation. Threats to both the designs are listed in **Table 13**.

In a *centralized* design, the data providers publish their data assets on to the platform transferring the data sovereignty over to the data marketplace. The data is stored by the platform and is transferred to the data consumer when the data is purchased. This involves the requirement of infrastructure for the storage of large volumes of data (Big data). Though it is implementation dependent, the big data storage is carried out with the help of data stores powered by flash storage supported by big data tools like Hadoop, Cassandra, NoSQL et cetera. These data stores are prone to threats because of the valuable commercial data they house. Because it is assumed to be a data store, the threat could not be one specific attack, rather could be mentioned as hacking comprising all the seven steps of a generic cyberattack:

Reconnaissance, Scanning, Access & Escalation, Exfiltration, Sustainment, Assault and Obfuscation (D. A. Smith, 2017). A successful attack at different stages of hacking causes damage to the data store. With respect to the assets they house i.e. data, a data breach causing the disclosure of proprietary data products published by providers on the platform could cause fatal damage to the data marketplaces in the form of financial, reputational and customer losses. If the data involved consists of the personal data collected from the users of the services provided by the data providers, the data breach can cause the violation of soft privacy leading to regulatory impacts on the data marketplace. Soft privacy refers to the violation of the privacy by an entity whose holds the personal data which is bought from other companies who directly collect from the users. The security techniques to safeguard the data on the data store could involve storing the data in the encrypted form. Furthermore, the servers need to be secured with firewall, anti-malware, intrusion prevention systems and system monitoring which form the basic infrastructure for security in organizations.

In a *decentralized* design, the metadata repository is the main asset managed by the data marketplace as part of the data inventory component. The reason being the data which is sold over the data marketplace are managed and maintained by the data providers themselves and provide only metadata information of the data sets to the marketplace. The metadata information is managed by the data marketplace and uses it in its broker service to connect the supply and demand. Further, a communication channel could be set up between the transacting parties to transfer the data being purchased on the marketplace. This aspect is part of data exchange service which will be dealt in the next subsection. The metadata management could involve database management and applications run on the server as supporting assets which could be subjected to attacks like Injection or malware to disrupt the metadata management. this could cause the disclosure of metadata information result in the loss of proprietary information. With this, the customer might lose the valuable resource and could hold data marketplace legally liable. The injection attacks could be overcome by using stored procedures over letting the customers query the metadata. These threats could also apply to centralized design as it also deals with metadata management along with data storage.

4.1.5 Data Exchange Service

This component merely signifies the transfer of the data from the data provider to the data consumer. The 2 designs (centralized and decentralized) apply here too. But in either of the designs, the threats remain the same as the core operation is the same: the transfer of large volumes of data through communication channel. In a centralized design, the communication channel between the data provider and the data marketplace; and between the data marketplace and the data consumer is the supporting asset. In the case of decentralized design, the communication channel set up between the data actors after they are matched on the data marketplace platform is the supporting asset. The threats to this supporting asset could involve the generic threats to the communication channel like eavesdropping, man-in-the-middle attacks as described in **Table 14**. A compromise in this area is very fatal for the data marketplaces as large volumes of commercial proprietary data are being transferred in this component. A data breach here could have the same impact as we discussed in the previous component resulting in violations of privacy agreements, loss of business-specific confidential data and so on. These threats could be mitigated by adopting a more sophisticated and secure mechanism to transfer the data between the parties. Common encryption methods could also pose risk since the resource involved is a significant one. More than just encryption, the business process of how the data assets are handled could be designed in a secure way with sophisticated security technologies.

4.1.6 Data Analysis Service

The business function assumed for this component in our architecture is like that of an app store, in addition to the data marketplace providing its own data analytics tools. It could allow third parties to upload their big data analytics tools and offer them to the customers of the data marketplaces. In this setting, the threats we could think of are with respect to the authenticity of the third-party data analytics tools. The tools could be uploaded by malicious third parties and hence, the tools can contain malicious constituents.

Table 13: Threats: Data Inventory

Primary Asset	*Supporting Asset*	*Threat*	*CIA violated?*	*Business Consequence*	*Mitigation Technique*
Centralized Design: Data assets published by the data providers; 2 variants: proprietary data and metadata	Data Store with flash storage coupled with servers powered by Hadoop, Cassandra, NoSQL et cetera.	Hacking • Reconnaissance • Scanning • Access & Escalation • Exfiltration • Sustainment • Assault • Obfuscation	-CIA of the data sets - Integrity of the DMP service	• Data Breach causing the disclosure of proprietary data of providers to attackers causing financial, regulatory and reputational losses • Soft Privacy violation in case of private data.	• Encryption • Firewall • Anti-Malware • Intrusion Prevention System System Monitoring
Decentralized Design: Metadata repository of the data products, metadata contains terms of usage	Database Management of metadata information	• Injection Attacks • Malware	- CIA of metadata - Integrity of the DMP service	• Disruption of the metadata management • Disclosure of metadata information of datasets of customers revealing metadata information which can be proprietary, contractual information etc.	• Stored Procedures • Encryption • Anti-Malware

Table 14:Threats: Data Exchange Service

Primary Asset	*Supporting Asset*	*Threat*	*CIA violated?*	*Business Consequence*	*Mitigation Technique*
Data being transacted / Data transfer mechanism	Communication channel	• Eavesdropping/ Traffic Analysis • Man-in-the-Middle • Malware	Confidentiality of the data; Integrity of the transfer service.	• Data Breach causing the disclosure of proprietary data of providers to attackers causing financial, regulatory and reputational losses. • Soft Privacy violation in case of private data.	Encryption

Table 15: Threats: Data Analysis Service

Primary Asset	*Supporting Asset*	*Threat*	*CIA violated?*	*Business Consequence*	*Mitigation Technique*
Data Analytic Tools: either downloadable or provided as SAAS	Third party analytics tools uploaded on marketplace similar to app store.	• Faulty Software • Malicious software uploaded by a malicious third party.	Integrity of the app store service of the data marketplace	Reputation loss and Legal liability for providing customers with malicious or faulty analytics tools	Screening and quality check of the analytics tools published by the third parties.

This could cause damage to the data sets subjected to analysis by the said tools resulting in a damage to the customer and in turn to the data marketplace in terms of legal liability and reputational deterioration. We could mimic an actual app store approach to overcome this threat by incorporating quality and security checks to the tools being provisioned by third parties. This way the customers could judge the authenticity of the services and trust the data marketplace. The above discussion is represented in ***Table 15***.The threat model satisfies our requirements in the sense that the threats are described at high-level to the high-level business functions of the data marketplace platform. Furthermore, the combination of the threats and business consequences of all the business functions associated with every functional component of the *Pre-MPC Data Marketplace Platform 1.0* represents the high-level threat landscape of the data marketplace platform. Furthermore, the high-level threat model signifies the *Pre-MPC Threat Model 1.0*

Although a pioneer effort towards exploring the threat landscape of the data marketplace platforms, because of the limitation of the *HLTM framework*, our work in this chapter only represents a ***baseline overview of the threat landscape of the data marketplace platform***. For this, reason, the threat model was subjected to validation later in Chapter 6 to obtain a more valid overview representing the actual threat landscape of the data marketplace platforms as perceived by actors from practice.

This page has been intentionally left blank.

5 Implications of MPC on Data Marketplaces

In this chapter, we first explore the scope of MPC technology utilized in the Safe-DEED project. These processes are then applied to the high-level architecture from Chapter 2 to create the *Post-MPC Data Marketplace Platform 1.0*. Following this, the effect of MPC incorporation on the threat model from Chapter 4 is explored.

5.1 MPC Technology in Safe-DEED

Secure computation is the solution for the famous problem called "*Two Millionaires problem*" where 2 millionaires wish to know who is richer without disclosing information about each other's wealth. Yao (1982) designed a protocol which solves this problem and it does so with the help of secure computation. The same solution has been researched to include more parties such that multiple parties can compute functions on the union of their data to produce desirable output without having to merge the individuals' actual data (Goldreich, 1998). This functionality finds an application in the context of data market where data security is a crucial aspect.

Multi-Party Computation (MPC) is a type of cryptographic protocols which allow functions to be computed over distinct datasets without having to share the data itself. As a result, the required knowledge from the data can be extracted without revealing the actual data. This characteristic is appealing to the data owners to create value as with MPC, they can share the business intelligence of their data without giving access to actual data. Several MPC methods already exist which carry out the above-mentioned functionality with mathematical sophistication. However, they suffer from scalability and performance limitations which restrict their usage in real-world applications.

Safe-DEED claims to overcome these limitations and provided a practical solution which will be tested with pilot cases. Safe-DEED claims to develop faster MPC protocols viable also for larger data sets by improving the computational and communication complexity of the underlying technical components.

To perform computation on the datasets using MPC protocols, it is necessary to know the function beforehand that needs to be applied on the data. The function signifies the knowledge that needs to be extracted from the data. Based on this function, the corresponding MPC protocol which can perform this function can be designed by selecting appropriate technical components. For example, if multiple companies want to perform mean and variance on some of their customer's data, then the functions, mean and variance need to be represented as circuit using addition and multiplication gates.

These addition and multiplication gates constitute the technical component blocks for building the MPC protocol. To help this cause, Safe-DEED proposes to develop those technical components required to execute different protocol of different functions. These technical components are referred as ***Safe-DEED Primitives***. These consist of convenient and easy-to-use methods to build protocols for the required function without requiring the deep understanding of the underlying technical aspects. These primitives involve cryptographic building blocks like low multiplicative complexity symmetric-key, garbled circuits, oblivious transfer and so on, which will be selected according to the requirements in designing the protocol.

The designed protocols need to support communication and hence, Safe-DEED also provides a network component powered by libraries such as OpenSSL or GnuTLS, which they refer as ***Safe-DEED Network.*** The whole offering of Safe-DEED comprising of the constituents, *Safe-DEED Primitives* and *Safe-DEED Network* is referred as ***Safe-DEED Component*** (Lupu, 2018) and is as illustrated in the schematic diagram in **Figure 8**.

Safe-DEED Component acts as a black box accepting the specification of the function and the data; and generates computational result which reflects the required outcome expected from the union of the data of the parties involved. Basically, Safe-DEED wishes to simplify the design of MPC protocols where the user who adopts the *Safe-DEED Component* only needs to decide on the function to be evaluated with other parties and has to supply the input data. Further, Safe-DEED takes care of the underlying technology in designing the protocol with the appropriate technical blocks.

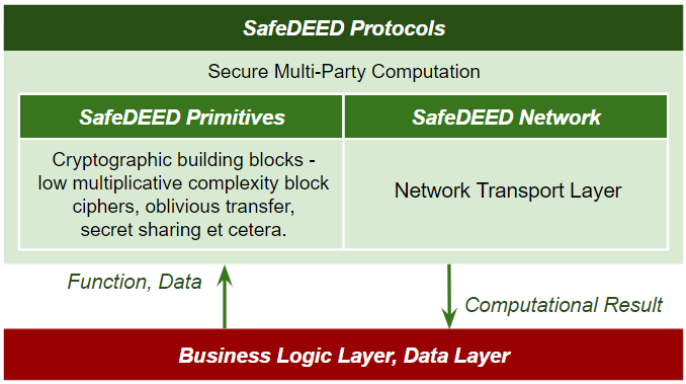

Figure 8: **Safe-DEED Component** for MPC Technology
Source: Lupu (2018)

5.1.1 MPC processes proposed by Safe-DEED

In Safe-DEED, we mostly focus on interactive approaches to MPC, i.e., where the parties involved have their data available simultaneously with all the actors for the computation to happen, i.e. in a synchronous way. This kind of process is represented in the schematic diagram in **Figure 9**.

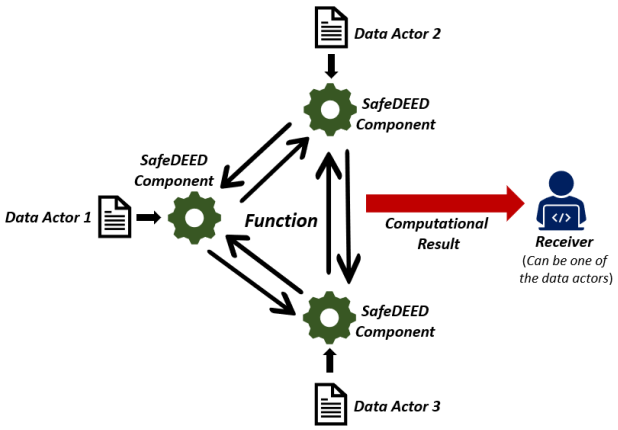

Figure 9: Interactive MPC Process

Safe-DEED also explores non-interactive approaches where data sharing can happen asynchronous. Safe-DEED proposes homomorphic encryption to enable this. Homomorphic encryption allows one to evaluate functions on encrypted data. Safe-DEED proposes a case where data providers encrypt their data to a dedicated receiver and send it to a dedicated aggregator who then evaluates the function on the cipher texts and forwards the computational result to the receiver. This kind of process is referred to as *multi-user data aggregation scheme*, see **Figure 10**. In this way, the process provides a non-interactive approach for data sharing which enables the providers to share data in an encrypted form which can be used later by the dedicated actors without having demand the presence of the data provider.

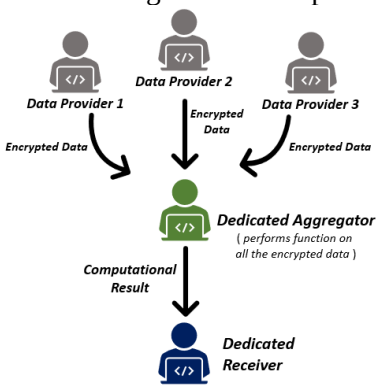

Figure 10: Non-Interactive MPC Process

These two approaches are supported by the literature dealing with the application of MPC and Homomorphic Encryption in the data marketplaces. Roman & Stefano (2016) designed a concept, *Trusted Data Marketplace* operating solely for the application of credit scoring. They design a reference architecture for a data marketplace platform where the actors involved in credit scoring can trade their data among each other. They suggest homomorphic encryption and multi-party computation as enabling technologies for the realization of their concept data marketplace where the physical data either remains with the data owner or is in encrypted form (by Homomorphic Encryption) stored on a cloud. They discuss 2 settings of data mining powered by MPC.

- In the first scenario, the data is held by 2 or more different parties and the data mining algorithm is run on the union of these parties' databases without letting each other know of other's data. This setting reflects the traditional MPC process where a function is computed on the union of databases from multiple parties to get a result.
- In the second scenario, some statistical data needs to be released for research or data mining. But the data might contain private information, hence, the data is modified first perhaps with anonymization so that the privacy is not compromised and meaningful results can be obtained from the anonymized data. This is a special case of the first scenario where, the parties anonymize their data before lending it for the computation where the MPC protocol carries out the union and the function execution.

These two scenarios reflect the first out of the two processes suggested by Safe-DEED. However, both processes can be implemented within the *Safe-DEED Component* and this could be integrated as a component or a feature into the architecture. In this way, *Safe-DEED Component* provides a way of incorporating the MPC technology into the high-level architecture from Chapter 2.

5.2 MPC Incorporation into the Data Marketplace Platform

The concept of MPC protocols can be related as a mechanism of the transferring the knowledge within the data from the data provider to the data consumer (without transferring the actual physical data). Consequently, the Safe-DEED component can be viewed as a component which enables the process of data exchange and hence, Safe-DEED Component was integrated into the ***Data Exchange Service*** of the high-level architecture as its business process.

The incorporation essentially makes the data marketplace platform a purely decentralized one as no physical data transfer is involved. Essentially, the platform will be responsible just for connecting the data providers, data aggregators and data consumers. Following the establishment of the relationship between the actors over the platform and the *Data Exchange Service* powered by *Safe-DEED Component* would be set up by the marketplace ad-hoc between the dedicated data actors outside the platform for them to interact with each other and share data. Furthermore, the computation of the function on the data from the involved actors will be performed by the *Safe-DEED Primitives (MPC Protocol)* according to the requirement. The computational result is then presented to the dedicated receiver through the communicational channel powered by *Safe-DEED Network*.

Furthermore, with an MPC enabled distributed data market place there would be limited need for a ***Data Inventory*** within in the architecture as the platform is decentralized now. So, the component gets transformed into just *Metadata Inventory* which just stores the metadata of the data provisioned to be transacted over the platform and will be used by the *Broker Service* which showcases the metadata to the customers through its functions. The backend features of *Broker Service* component also go through changes where the management activities like cataloguing and curation activities are done only for the metadata of the data. Since there will be no data publishing on the platform, the data aggregator steps out of the umbrella of data providers. The aggregator's function with respect to this design is aggregation of the data and not publishing the aggregated data. Hence, the data aggregator becomes a distinct actor who will avail the platform to provide his aggregating services. Meanwhile, the data provider actor is transformed into just data owner who holds the different types of data like raw data, polished data, formatted data et cetera and provisions the data on the platform by publishing its metadata. The actors, *Data Collector* and *Data Manager* considered earlier now fall under *Data Owner* as they own and offer data on the platform.

With respect to the functional requirements, *Secure Data Exchange* requirement is enabled by *Safe-DEED component* with its MPC protocols. *Data Sovereignty* is retained by the data provider as the provider holds the control over his physical data. *Data Governance* is also taken care of by the data provider as he becomes responsible for the management and maintenance of his data. The modified high-level architecture of the data marketplace platform after the incorporation of the *MPC technology*, signifying the *Post-MPC Data Marketplace Platform 1.0* is illustrated in **Figure 11**, with the modified elements highlighted in ***yellow***. The *Data Exchange Service* is depicted separately in **Figure 12** which reflects the functioning of the *Safe-DEED Component.*

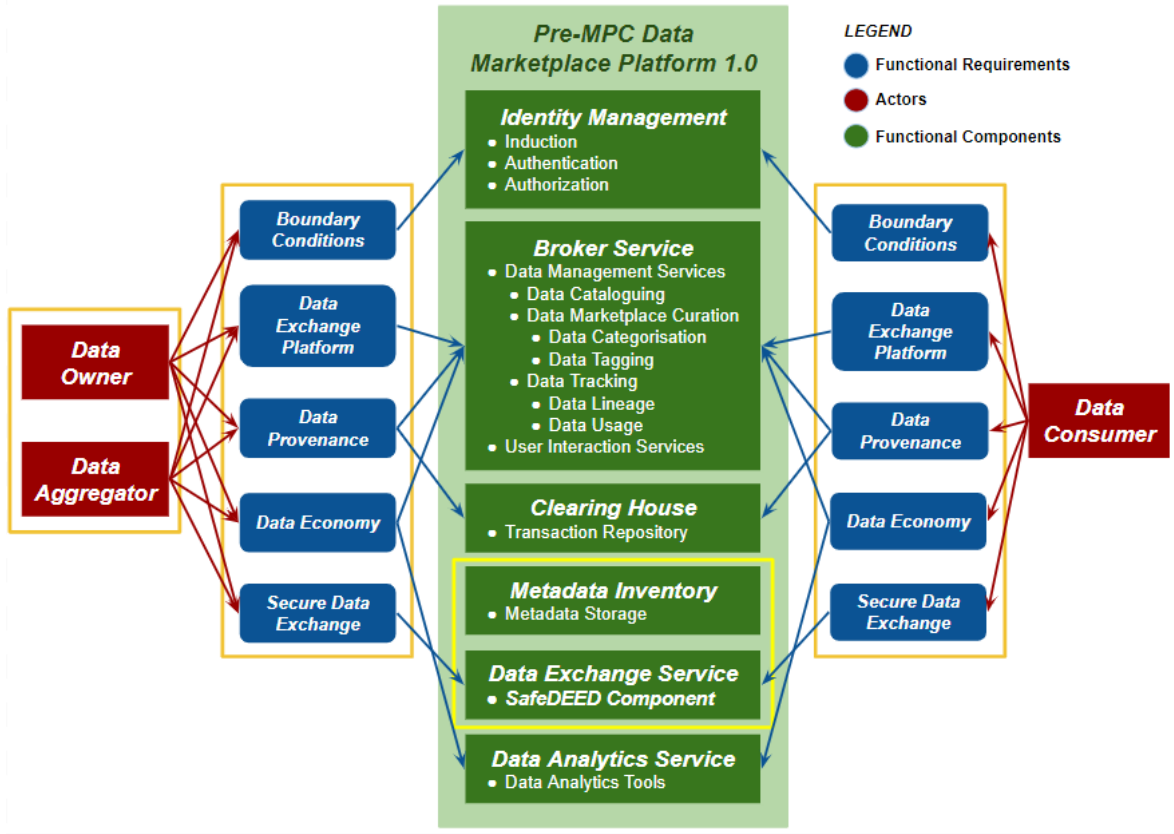

Figure 11: Post-MPC Data Marketplace Platform 1.0

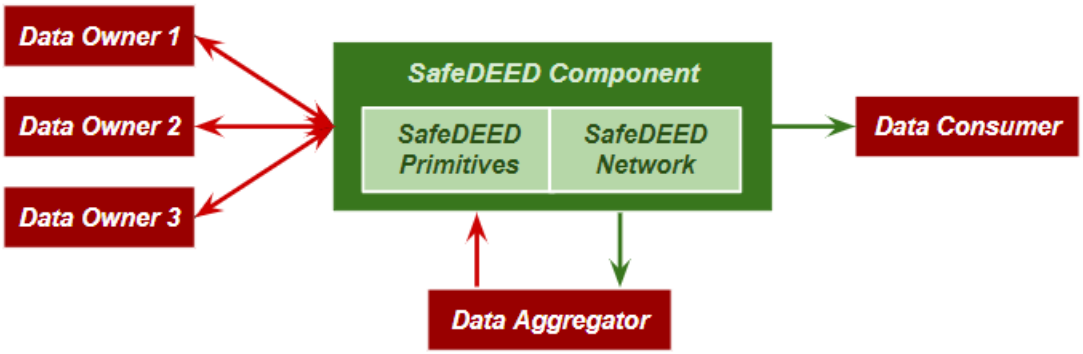

Figure 12: Data Exchange Service enabled by Safe-DEED Component powered by MPC

5.3 Effect of MPC Incorporation on the Threat Model

The functional components that undergo major change with the incorporation of MPC technology are Data Inventory, which is now, *Metadata Inventory* and *Data Exchange Service*. As a result, the incorporation affects the threats of only these two components and not that of any other component in the architecture.

5.3.1 Post-MPC Threats: Metadata Inventory

Since the platform now is decentralized, the commercial proprietary data stays at the site of the data owner and there is no transfer of physical data over the platform, the incorporation overcomes the risk of data breach or the violation of privacy (in case of private data). The platform now houses only the metadata of the data provisioned by data owners. The threats identified in the threat model still apply to this metadata. However, the risk with the disclosure of the metadata is less compared to the disclosure of the commercial data. This way, the risk involved with the inventory is reduced by the incorporation of MPC technology in the data marketplace platform. The threats associated with the modified *Metadata Inventory* component is listed in **Table 16**.

Table 16: Post-MPC Threats: Metadata Inventory

Primary Asset	*Supporting Asset*	*Threat*	*CIA violated?*	*Business Consequence*	*Mitigation Technique*
Metadata repository of the data products, metadata contains terms of usage	Database Management of metadata information	• Injection Attacks • Malware	• CIA of metadata • Integrity of the DMP service	Disruption of the metadata management, disclosure of metadata information of datasets of customers revealing metadata information which can be proprietary, contractual information etc.	• Stored Procedures • Encryption • Anti-Malware

5.3.2 Post-MPC Threats: Data Exchange Service

Since the data exchange now happens via the *Safe-DEED Component* using an MPC Protocol, the threats leading to a data breach, impose a lesser risk. Information in the communication channel is an intermediate result obtained during the protocol execution but not the actual data. So, when the communication channel is compromised by an outsider's attack, the breached information will not be of any use to the attacker as the physical data is not there. However, the threats causing the breach of the communication channel disrupts the data marketplace service, compromising its integrity. The threats associated with the modified *Data Exchange Service* are listed in **Table 17**.

Table 17: Post-MPC Threats: Data Exchange Service

Primary Asset	*Supporting Asset*	*Threat*	*CIA violated?*	*Business Consequence*	*Mitigation Technique*
• ***Data being transacted*** • ***Data transfer mechanism***	Communication channel powered by ***Safe-DEED Component***	• Eavesdropping Traffic Analysis • Man-in-the-Middle • Malware	Integrity of the data transfer service and in turn, integrity of the DMP service.	Service disruption of DMP	Intrusion Prevention system.

5.4 Summary

In this chapter, we discussed how MPC technology changes the architecture and threats associated with a data marketplace platform. We first illustrated the conceptualization of MPC technology in Safe-DEED, and how the project intends to materialize its process. It was deduced that Safe-DEED materializes MPC technology with its ***Safe-DEED Component*** comprising of the ***Safe-DEED Primitives,*** which provides the technical blocks required for building the protocol and ***Safe-DEED***

Network, which provides a communication channel for the execution of the protocol. This *Safe-DEED Component* provides a black box way of incorporating MPC technology for the customers who could just choose the required function and provision and let the Safe-DEED Component to build and execute the protocol. The *Safe-DEED Component* was integrated into the *Data Exchange Service* as they both represented a mechanism of transferring data or the knowledge inside it from the data owner to the data consumer. As a result, the platform would become decentralized where the actors can meet over the platform and the *Data Exchange Service* enabled by the *Safe-DEED Component* is set up ad-hoc by the marketplace, outside the platform, when the actors execute the protocol and share data. This move also eliminated the need for *Data Inventory* which now is transformed into Metadata Inventory which stores and maintains metadata of the data provisioned on the platform. Furthermore, the requirements of *secure data exchange* are reinforced; while *data governance* and *data sovereignty* are moved to the site of the actor owing the decentralized transformation of the platform. Furthermore, there is a change in the way the customers are represented and now they are comprised of Data Owners, Data Aggregators and Data Consumers. This collectively is the effect of MPC incorporation into the high-architecture of the data marketplace platform. The resulting updated architecture represents the *Post-MPC Data Marketplace Platform 1.0.*

The effect of this MPC incorporation on the threat model from Chapter 4 is that this move minimizes the risks associated with the components, *data inventory* and *data exchange service* as the element of physical data is eliminated from the components. Apart from these components, MPC does not interfere with the threats of rest of the components. *Ultimately*, MPC technology increases the security value of the data marketplace platform by addressing the most significant factor, data handling on the data marketplace platform in a *Security-by-Design* way. The resulting refined threat model represents the *Post-MPC Threat Model 1.0*.

6 Model Validation

In this chapter, we validate the organization and threat models developed during the *conceptualization* phase to generate more valid artefacts and valid concepts. With this agenda, a qualitative study was conducted by interviewing experts in three subject areas: *data marketplaces, threat modelling* and *MPC technology*. Our research methodology was formulated by establishing its different parameters: *Design, Participants, Procedure* and *Analysis* as suggested by Kraus, Fiebig, Miruchna, Moller, & Shabtai (2015).

6.1 Methodology

The research strategy generally employed by researchers for theory development is *Grounded Theory.* Grounded Theory is a strategy to derive a theory inductively from the data (Corbin & Strauss, 1990). The process involves generating a theory by collecting the data, analyzing the data which directs what data to collect next until a saturation is reached; finally, to end up with an inductively derived theory. In Grounded Theory, the theory is derived solely from the collected data. Hence, Grounded Theory can be an extreme way which truly builds a theory. However, there is a less extreme variant of Grounded Theory called, ***Middle Ground Approach*** which refines an already existing theory (Sekaran & Bougie, 2013; d9). This method necessitates an initial list of codes and categories informed by an already existing theory which directs both the data collection and then, the data analysis process. This approach is a perfect fit for our research agenda of validating the artefacts from the conceptualization phase. Hence, we adopted a *Middle-Ground Approach* for our research. The first iterations of the four conceptual models from the conceptualization phase constituted the *initial list of codes and categories* which also directed the design of the interview questions and thereby, the data collection. The process of the data analysis remains the same which involves constant comparison of newly collected data with the existing list of categories and codes and then updating the theory to reflect the insights from all the collected data until theoretical saturation is reached.
The initial setup of the Middle Ground Approach i.e. explicating the initial set of categories and codes is performed first before getting into the actual methodology. We focus on ***three subject areas (SA)***:

- SA1: Data marketplaces
- SA2: Threat modelling for data marketplaces
- SA3: Multi-Party Computation (MPC) for data marketplaces

Related to these subject areas, ***four research foci (RF)*** were formulated which signify the validation agendas for the *Artefacts 1.0* of the conceptualization phase:

- RF1: Validate *Pre-MPC Data Marketplace Platform 1.0* to generate *Pre-MPC Data Marketplace Platform 2.0*
- RF2: Validate and refine the concept of MPC Incorporation into the data marketplace platform and further, generate *Post-MPC Data Marketplace Platform 2.0*
- RF3: Validate *Pre-MPC Threat Model 1.0* and generate *Pre-MPC Threat Model 2.0*
- RF4: Identify the effect of MPC incorporation on the threats from the *Pre-MPC Threat Model 2.0* and further, generate *Post-MPC Threat Model 2.0*

Based on this, we identified 10 topics:

- T1: The data marketplace platform designs (*RF1*)
- T2: The functional requirements of the data marketplace platform (*RF1*)
- T3: The customers of the data marketplace platform (*RF1*)
- T4: The functional components of the data marketplace platform (*RF1*)
- T5: The HLA framework (*RF1*)
- T6: The perception of MPC technology from conceptualisation phase (*RF2*)
- T7: The MPC incorporation into the data marketplace platform (*RF2*)
- T8: The HLTM framework (*RF3*)

- T9: The threat landscape of the data marketplaces *(reflected by the threats and business consequences in Pre-MPC Threat Model 1.0)* (*RF3*)
- T10: The effect of MPC incorporation on the threat landscape of the data marketplace platform (*which is the validation of Post-MPC Threat Model 1.0*) (*RF4*)

We validated each of these ten topics using the *Middle-Ground approach*. In each topic, we dealt with several ***theoretical concepts*** which were used to answer the corresponding sub-research questions of the conceptualization phase. These concepts comprised of *definitions, interpretations, descriptions, taxonomies, architectures, frameworks, threat models, processes* et cetera. Basically, these included every concept associated with the resulting artefacts from the conceptualization phase. These theoretical *concepts* and their corresponding low-level information in each topic constituted the ***initial list of categories (C)*** and ***codes (C*)*** for that topic's validation activity. We derived ten sets of initial lists of categories and codes for the 10 topics and their corresponding theoretical concepts and these ten lists collectively represent the initial specification of categories and codes required for the Middle-Ground Approach methodology. This prerequisite information formulated prior to starting the validation phase is illustrated in the form of a hierarchy in **Figure 13**. Furthermore, we list the topics associated with each research focus mapped to their respective subjective areas in **Table 18**.

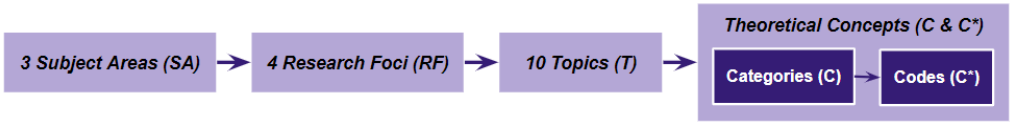

Figure 13: Initial Specification for Middle-Ground Approach

Table 18: Subject Areas, Research Foci and Topics

Subject Area	*Research Focus*	*Topic*
SA1: Data Marketplaces	***RF1**:* to validate *Pre-MPC Data Marketplace Platform 1.0* and generate *Pre-MPC Data Marketplace Platform 2.0*	***T1:** Data Marketplace Platform Designs*
		***T2:** Functional Requirements of the Data Marketplace Platform*
		***T3:** Customers of the Data Marketplace Platform*
		***T4:** Functional Components of the Data Marketplace Platform*
		***T5** HLA Framework*
SA1: Data Marketplaces & SA3: MPC Technology	***RF2**:* to validate and refine the concept of MPC Incorporation into the data marketplace platform and further, generate *Post-MPC Data Marketplace Platform 2.0*	***T6:** Perception of MPC Technology*
		***T7:** MPC Incorporation into the Data Marketplace Platform*
SA1: Data Marketplaces & SA2: Threat Modelling	***RF3**:* to validate *Pre-MPC Threat Model 1.0* and generate *Pre-MPC Threat Model 2.0*	***T8:** HLTM framework*
		***T9:** Threat Landscape of the Data Marketplaces*
SA1: Data Marketplaces, SA2: Threat Modelling & SA3: MPC Technology	***RF4**:* to deduce the effect of MPC incorporation on the threats from the *Pre-MPC Threat Model 2.0*	***T10:** Effect of MPC Incorporation on the Threat Landscape of the Data Marketplace Platform*

6.1.1 Expert Interviews

The research method for collecting qualitative data was chosen to be ***Expert Interviews***. Interviews are one of the primary qualitative data collection methods which is widely used to collect rich data for exploratory studies in general business settings (Sekaran & Bougie, 2013). Expert interviews are a specific kind of interviews where subject area experts are specifically interviewed to obtain expert knowledge about the focal subject area. Given that the research foci of our objective are related to new subject areas of which knowledge is not out there yet, we have adopted expert interviews to be our qualitative data collection method as only experts can provide insights regarding these new subject areas. Regarding the type of interviews, it was decided to use ***Semi-Structured Skype Interviews***. Semi-structured interviews are the ones with some pre-defined open ended questions in an order which helps in establishing the focus on a subject while giving the flexibility to explore deeper into the subject through a follow-up discussion for the questions (de Reuver, 2019). Since our purpose of doing

qualitative data analysis is to validate the concepts and artefacts from the conceptualization phase and to update them with deeper insight, we adopted the semi-structured approach for the interview protocol. The interview questions were prepared by basing the questions on the respective concepts present in the initial list categories and codes associated with each topic's exploratory study. This way the initial list of categories and codes served their purpose in the Middle Ground approach which is to direct the data collection activity; in this case, interviews. The questions helped to explore each concept deeper while clarifying sketchy insights with follow-up questions; most of the times turning the interview into brainstorming session on the focal subject area.

6.1.2 Participants & Sampling

We carried out *judgement sampling* to choose the participants as it fit our objective of obtaining expert knowledge on the subject areas. Judgement sampling is a variant of purposive sampling which is used when specialized information is necessary for the study which is not available easily as that information is not mainstream (Sekaran & Bougie, 2013). The experts in our three subject areas: *Data Marketplaces*, *Threat Modelling* and *MPC technology* were considered for the interviews. The profiles of each subject area expert were formulated as follows,

- ***Data Marketplaces*:** Researchers working in the field of data intermediaries, data exchange mechanism and data marketplaces
- ***Threat Modelling***: Researchers and industry experts working in the cybersecurity domain
- ***MPC technology***: Researchers working in the Safe-DEED: Safe Data Enabled Economic Development project who are conceptualizing and developing the MPC technology.

Experts were approached via the Safe-DEED project's network. The prospects were invited for the interviews with email invitations informing the experts beforehand the kind of work being dealt and what was expected of them; before they accepted the invitation. Since the purpose of the interviews was to validate the artefacts from the conceptualization phase, it was necessary to familiarize the experts with the concepts associated with the relevant topics beforehand so that they would have better context and understanding of the concepts before getting into the interview; thereby potentially increasing the chances of their answers to be more informed and nuanced. For this purpose, the descriptions of the artefacts (as relevant for each prospective expert's subject area) consisting of the concepts were compiled into a document and was sent as an attachment with the email invitation to the respective subject area experts. Out of ten invited prospects, the experts who responded and were eventually interviewed are listed in Table 20 along with their relevance to our research.

Table 19: Experts interviewed for the Validation Phase

Expert (E)	*Role*	*Relevance*
E1	Post Doc researcher working on a project on the conceptualizing of a data marketplace in the airline industry.	Expertise in data exchange mechanisms and data marketplaces.
E2	Research Coordinator of Safe-DEED. Working closely with research partners to develop the enabling technologies for B2B data sharing like MPC, Data Valuation etc. Also working closely with Data Market Austria in its conceptualization.	Experience in materializing a real-life data marketplace, Data Market Austria.
E3	Manager in the domain of Cybersecurity and Privacy at a major consulting firm. Provides auditing and security assessment services to business clients.	Expertise in threat assessment and security frameworks.
E4	Researcher working on the implementation of Multi-Party Computation (MPC).	Expertise in MPC technology and its applications.

6.1.3 Procedure

The expert interviews were semi-structured interviews and were conducted via video chat. Prior to the interview, the experts were directed to be familiar with the concepts described in the attached document and were asked to have a copy of the same document with them so that it is easier for them to follow when the concepts are referred during the interview. Before starting the interview, the consent of the expert was taken verbally to record, transcribe and use the insights from the interview in our research. After taking the consent, the interviews were recorded. Once the recording started, the same consent

was taken verbally again so that the consent was also on record. After this, it was asked to confirm if the expert had a chance to familiarize himself/herself with the concepts of the relevant artefacts prior to the interview. An overview of the research and the relevant artefacts was verbally described for almost 10 minutes before starting the actual interview, to ensure that experts who may only have skimmed the document are also familiar with the concepts. This solution was further solidified by verbally explaining each concept being dealt before asking the corresponding question. The interview was carried out by asking the previously-prepared semi-structured questions, the follow-up questions and the follow-up discussion which went on until a comprehensive understanding was reached on each concept. **Table 20** shows the topics on which the insights were provided by each expert.

Table 20: Topics validated by each Expert

Topic/Expert	*E1*	*E2*	*E3*	*E4*
T1: Data Marketplace Platform Designs	✓	✓	-	✓
T2: Functional Requirements of the Data Marketplace Platform	✓	✓	-	-
T3: Customers of the Data Marketplace Platform	✓	✓	-	-
T4: Functional Components of the Data Marketplace Platform	✓	✓	-	-
T5: HLA Framework	✓	-	-	-
T6: Perception of MPC Technology	✓	✓		✓
T7: MPC Incorporation into the Data Marketplace Platform	✓	✓		✓
T8: HLTM framework.	✓	-	✓	-
T9: Threat Landscape of the Data Marketplaces	✓	-	✓	✓
T10: Effect of MPC Incorporation on the Threat Landscape of the Data Marketplace Platform	✓	-	-	✓

6.1.4 Analysis

Each interview was transcribed, and analyzed right after the interviews to refine the concepts further. This allows us to use the refined concepts in the further interviews. This helped in deepening the understanding of the concepts as the number of the interviews progressed. However, the formal qualitative data analysis was carried out after all interviews were done. The qualitative data analysis was carried out with the traditional steps: *Data Reduction, Data Display* and *Drawing Conclusions*.

In *Data Reduction*, since we already had the initial list of categories and codes of each topic, we moved directly to the second phase of coding, ***Analysis phase: Axial Coding*** (de Reuver, 2019b). Here, we mapped the statements and insights from the interview transcripts to their appropriate categories and codes. Subsequently, with this mapping, we analyzed and carried out the refinement, updating and modification of the concepts of all the categories and codes. After this process, with the data that is left unrelated to the existing codes, new codes were created for these unmapped insights and were assigned to their appropriate categories and topics. The whole data reduction was done manually using a data log book where we documented the constant comparison between the interview transcripts and the then list of categories and codes. No software was used to carry out the data reduction. As a result, there was no illustrative way to visualize the data reduction and hence, data reduction was decided to be represented in a qualitative way (basically, in ***words***) as opposed to the traditional ways of data visualization (like matrix, timeline, networks, actor network, process (de Reuver, 2019b)). However, we illustrate the categories and codes in either of the lists (initial and updated) are illustrated before and after the analysis in each topic in the form of *figures, tables, lists, hierarchies* or just *textual descriptions*

Moving on, the ***Results were Analyzed*** for each topic signifying the *Data Reduction & Data Display* step of qualitative data analysis. Here, the data mapped to the appropriate concepts i.e. the data reduction and data display are *represented in a qualitative way* by relating it to the respective expert. If the resulting code relates to the concepts already associated with the initial set of categories and codes, they are represented in *Italic* font while the newly emerged concepts and their codes are displayed in ***bold-face*** font; both contributed towards generating the updated list of categories and codes. Following this, we started ***Drawing Conclusions*** for each topic signifying the last step of the same name of the qualitative data analysis. The corresponding sections collectively contain the updated iterations of all concepts refined, updated or modified after incorporating either the quoted insights, further analyses or further implications to obtain a more valid concept for each topic. These represent the updated list of

categories and codes associated with the concepts of each topic. There was an anomaly with one of the topics, *T9* for which all the initial list of categories and codes were discarded during the analysis process. We will hence revisit it in the discussion in Chapter 7. Later, we generated a new list of categories and codes from the interview transcripts alone by carrying out the first phase of data reduction which is, ***Exploration phase: Open Coding*** (de Reuver, 2019b)**.**

6.2 Validation of Pre-MPC Data Marketplace Platform 1.0

The topics and corresponding theoretical concepts associated with the research focus, *RF1* are validated here. The artefact under consideration here is the *Pre-MPC Data Marketplace Platform 1.0* built in Chapter 2. The following 6 topics under *RF1* are validated in the upcoming subsections.

- ***T1:*** *Data Marketplace Platform Designs*
- ***T2*****:** *Functional Requirements of the Data Marketplace Platform*
- ***T3***: *Customers of the Data Marketplace Platform*
- ***T4*****:** *Functional Components of the Data Marketplace Platform*
- ***T5:*** *HLA Framework*

6.2.1 Data Marketplace Platform Designs

In Chapter 2, the potential platform designs of the data marketplaces were discussed as proposed by Koutroumpis, Leiponen, & Thomas (2017) which involved, ***centralized, decentralized*** and ***collective platforms***. However, these were predictive conceptualizations proposed based on the economic perspective of the institutional requirements: *boundary conditions*, *rules* and *monitoring mechanism*. In addition to the functional requirements, these conceptualizations do not consider other design aspects of data marketplaces like that architectural aspects, business processes, enabling technologies like homomorphic encryption, multi-party computation et cetera and their maturity to implement into the data marketplaces. The designs were just theoretical frameworks and hence, they do not reflect the real-life platform designs of the data marketplaces. For this reason, this topic, *T1: Data Marketplace Platform Designs* was considered for an exploratory study under the hope to enhance their understanding with expert insights.

The theoretical concepts associated with this topic, *were* analyzed by relating them to the insights of experts *E1* and *E2*. The initial list of codes in this topic derived from Chapter 2 were:

- ***T1:*** *Data Marketplace Platform Designs*
- ***C1:*** *Centralized Platform*
- ***C2:*** *Decentralized Platform*
- ***C3:*** *Collective Platform*

6.2.1.1 Results & Analysis

When asked about the real-life data marketplaces and their platform designs, *E1* responded by saying "*the term, data marketplaces, is a bit overused*" and suggested that even single domain data provider who provisions data over a cloud also calls himself a data marketplace. This insight is in similar lines with our criticism towards the systematic survey of Schomm et al. (2013) which includes even data vendors in their survey of data marketplaces and subsequently, suggests the focal data marketplaces (multilateral B2B data marketplace) in this research as just one of the categories in their classification. *E1* suggests that the ideal design of a data marketplace is to have a "*distributed system similar to Internet Exchange"* where anybody can hook up to the marketplace and carry out data exchange with anybody. We name this code as ***truly many-to-many data marketplaces.*** *E1* claims that this kind of design is theoretically possible and is being worked on. However, the execution of such a marketplace is complex and the idea is not realized yet owing to many reasons. Speaking on the real-life data marketplaces, *E1* suggested that the actual data marketplaces that do exist are formed in the lines of a ***consortium*** where "*parties within an industry come together to figure out a way to share data such that it is profitable for all the parties*" involved. Following this, the parties figure out a ***use-case*** to generate value out of data and create an architecture of a *data marketplace for that specific use-case* with fixed actors and fixed

processes**.** Furthermore, *E1* touches upon the possibility of *centralized* and *decentralized* data marketplaces in the same meaning as our initial codes; which is based on where the physical data resides. He says that *decentralized design* is operational with the help of a "*key management system*". In this case, a data provider holds the data and provisions his data with the help of public key encryption where the dedicated data consumer holds the private key and gets access to that data. Since this involves a requirement for governance to manage the public and private keys, this kind of model would not realize *truly many-to-many data marketplaces* where governance is complex because of its true many-to-many nature. However, in a closed consortium with fixed limited members, the governance of key management and subsequent data transactions is feasible. *E1* suggests another way of materializing *decentralized* design is by putting the data on ***block chain*** "*but it is not feasible yet for real-life application".*

When asked *E2* about the platform designs of data marketplaces, he reflects on a ***truly many-to-many data marketplace*** that it is not possible to realize it for various reasons. The *absence of data sharing culture* is one of them. *E2* suggests that in a practical sense, the realization of data marketplaces is driven by the ***use-case*** through which the data is utilized. Once the *use-case* is developed, data can be brought onto the platform easily from the data owner. However, *E2* also suggests that it is difficult to foresee a *use-case* without the availability of the data and its details. *E2* relates to this as a chicken-egg problem. However, *E2* also discusses the possibility of a platform where an innovator who innovates the use case can search for the appropriate data on that platform. On this kind of data marketplace platform, the innovator can also browse through the data catalogue using the metadata provided on the platform by the data owners and if interesting data is found, can innovate a *use-case*. *E2* reflects that the former case is more likely than the latter one. *E2* calls the latter kind of data marketplaces as ***"serendipity model"***. *E4* also echoes the *serendipity model* by referring it as a platform where the companies who have data and the companies who want to run statistics on such data can find each other.

6.2.1.2 Conclusions

Combining the above-discussed insights into the initial codes of *T1*, we built a taxonomy for the platform designs of the many-to-many data marketplace platforms reflecting the expert insights; thereby, replacing the previous classification. The taxonomy represents the updated list of categories and codes of *T1*. Broadly, the taxonomy consists of 2 categories of platform designs based on where the data resides: *Centralized* and *Decentralized.*

- In ***centralized*** design, the data is transferred from the data owner and stored on the platform and the data consumer finds the data on the platform and downloads it for his/her use. Since the owner loses the control over the data, only low value data like open data is transacted through such platforms.
- In ***decentralized*** platforms, the data resides at the data owner's site and is accessed only by dedicated data consumer or data aggregator through some encrypted channel. Since the data owner has the control over his data and the data consumer is allowed to access that data over contractual obligation facilitated by the platform, high value data can be transacted on such platforms. Further, in *decentralized* design, there can be 2 variants based on the design specification of the data marketplaces related to its ecosystem design, technological architecture design et cetera. The variants are *truly many-to-many data marketplace, block chain based data marketplace* and *closed consortium data marketplace.*
  - The ***truly many-to-many data marketplace*** is the ideal design where anybody can log into the platform and provision their data to anybody else on the platform, as suggested by *E1* and reflected by *E2*. This is the end goal for the species of data marketplaces which is feasible only in time when other factors like technological maturation, data sharing culture et cetera come together.
  - The species of the ***blockchain based data marketplace*** is straightforward as suggested by *E1* where the data transaction happens through a block chain. The data owner uploads his data to the blockchain and the data consumer access the data on the blockchain. Meanwhile, the blockchain monitors all the activities being carried out on that data which is stored, and any anomaly will be reported. This design relates to the *decentralized platform* as suggested by

Koutroumpis et al. (2017). The design is being worked upon and is expected to materialize once the blockchain technology attains mainstream maturity which is not very far in the future.

- The ***closed consortium data marketplace*** are the data marketplaces formed by parties within an industry to share data among each other. This variant is similar to the *collective platforms* as suggested by Koutroumpis et al. (2017) which already operate in the real world. Furthermore, in *closed consortium data marketplaces***,** we have included 2 more subcategories based on the business process associated with them. They are: *use-case based data marketplace* and *serendipity model data marketplace*.
  - In a ***use-case based data marketplace***, a fixed number of data actors come together to form an architecture driven by a specific use-case which defines the business process of the data marketplace. In this variant, the business process and the roles of the actors in the architecture will be fixed while the companies representing the actors can plug-in as and when necessary to transact the data, satisfying the many-to-many criteria. The data marketplace proposed by Roman & Stefano (2016) can be attributed as an example for this variant. This design was seconded by *E1* and *E2* as the most-likely and a realistic design for a data marketplace as this design practically exist in operation in the real world.
  - The ***serendipity model data marketplace*** is a platform where the data owners within the consortium can showcase their data in the form of metadata for the potential data consumers in need of that data and consequently, form a relationship and share data among only each other in an ad-hoc sort of way with a communication channel. Here, other data actors like data managers and data aggregators also showcase their services on the platform to find data partners. This design is more flexible with no business process fixed for the data trading but is formed when the data actors find each other with their data and corresponding use-case for the utilization of that data.

The taxonomy reflects the final list of the categories and codes of *T1* and is illustrated in **Figure 14** in the form of a hierarchy. This serves as an update to the classification of Koutroumpis et al. (2017) and also extends the category of ***Data Market Place*** in the classification of Schomm et al. (2013).

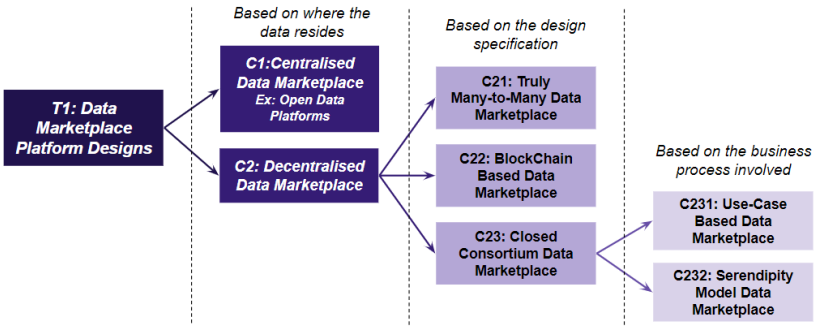

Figure 14: Data Marketplace Platform Designs Taxonomy 2.0

After all these different designs of data marketplace platforms were established, it was deduced that our focal data marketplace (multilateral B2B data marketplace) as illustrated using *Pre-MPC Data Marketplace Platform 1.0* from Chapter 2, related to the *Serendipity Model variant* in the *closed consortium* category from the taxonomy. We combined this insight and refined our scope. The resulting species of the data marketplaces which was focused from then on was ***Many-to-Many B2B Decentralized Serendipity Model data marketplaces.*** In the rest of the document, when we refer the term data marketplace, we mean this species. The reason for doing so was that through our knowledge from the study on data marketplace so far, it was deduced that this species represented the most generic form of a data marketplace which coincided with the one referred in *RT1*.

6.2.2 Functional Requirements of the Data Marketplace Platform

The functional requirements were discussed comprehensively in Chapter 2. However, the actual meaning of these requirements was needed to be understood to check if it reflects the same as our

interpretation. Hence, the topic, *T2: Functional Requirements of the Data Marketplace Platform* was included as an exploratory study to be validated and refined.
The theoretical concepts associated with this topic were analyzed by relating them to the insights of experts *E1* and *E2*. The initial list of categories and codes in this topic derived from Chapter 2 were:

- ***T2:*** *Functional Requirements of the Data Marketplace Platform*
- ***C1:*** *Boundary conditions*
- ***C2:*** *Data Provenance*
  - ***C21:*** *Data Lineage*
  - ***C22:*** *Change of Ownership of Data Point*
- ***C3:*** *Data Governance*
  - ***C31:*** *Management of Data*
  - ***C32:*** *Data Exchange Traceability*
  - ***C33:*** *Data Usage*
- ***C4:*** *Data Economy*
  - ***C41:*** Revenue
- ***C5:*** *Data Sovereignty*
  - ***C51:*** Handling Permissions
  - ***C52:*** Usage Restriction
  - ***C53:*** Data Contacts
- ***C6:*** *Secure Data Exchange*
- ***C7:*** *Data Exchange Platform*.

6.2.2.1 Results & Analysis

When discussing about the general requirements for a data marketplace platform, *E1* suggested that the starting point here is having a ***governance*** model. He expands on enforcing *governance* as an "*an authority who manages all the parties and activities*" on the data marketplace platform. One of the activities involves handling the ***legal aspects*** comprising of contracts which "*contain the terms of what can be shared with who, which data can be shared using which algorithm, what computing functions can be done in this algorithm, timeframes, quality of the data et cetera*". Furthermore, *E1* adds another requirement associated with the *governance* which is ***trust mechanism***. *E1* says *trust mechanism* is enforced again with independent authorities like ***Certification Authority***, ***Auditing Authority*** et cetera. These authorities with their activities bring about the trust on the data marketplaces in an indirect and intangible way. *E1* reflected on our assumption of enforcing data governance just through technology alone that technology can "*kind of enforce the governance but there is no way to restrict technologically when someone among the parties can just copy the data and run away with it*". *E1* says that something like this can only be tackled from the legal angle, with an authority and not from the technological angle. Basically, *E1* says "*the complex thing is to find a right coordination between the technology and legal aspects to have a complementary effect*". Basically, the requirements can only be enforced if both the aspects of technology and legal angle are in place and it cannot be done by just one of them. *E2* did not touch up on these issues and went right about reflecting on the functional requirements we had compiled from the literature.
Moving on to reflecting on the initial set of functional requirements from Chapter 2, *E1* and *E2* had several comments.

- ***Boundary conditions:*** This requirement is referred as necessary by *E1* and stated that it also depends on how it is implemented, and it is a topic in itself to explore. *E1* also suggested that this requirement is part of the ***governance*** aspect. *E2* reflects that our phrasing of *boundary conditions* is good and states that it "*is required*".
- ***Data Provenance:*** *E1* stated data provenance as "*an important aspect*" and suggested that this requirement is enforced through auditing of the transactions. The audit trail gives the *data provenance*. Both *E1* and *E2* had problem with our phrasing in the description of the concept of data provenance. *E1* suggested that the phrase "*change of ownership of data points*" used in the description of data provenance is not clear and it should be defined precisely given that it can have different implications. *E1* says technologically, the *ownership of data* can be defined in terms of *ownership of private key* to access the data in which case, the switching of private keys signifies the

change of ownership of data. A ***key management*** component will come in place there as part of the identity management. However, *E1* says there is risk involved here if it is done without any governance as in that case, even if the change of ownership of data, the owner can have a copy of the same data and he can sell it to other party. So, the governance model should take care of this aspect such that the ownership change happens according to the terms in the ***contract.*** *E2* also reflects on our phrasing of *change of ownership of data* and discards the concept saying that in data markets, ownership of data does not exist and what exists are ***licenses***. *E2* says that "*there is no process involved where the change of ownership happens".* Furthermore, *E2* clarifies the meaning of ***data lineage*** by saying *"it is the transformation of the data from its origin to the current state*" and ***data usage*** by relating it to "who has access to the data, who accessed it and whether they accessed it or not. These two concepts are the basis of *data provenance*.

- ***Data Governance:*** *E1* states *data governance* as "*the most important requirement which establishes the legitimacy of the data marketplace*". It is enforced through an authority actor who oversees all the operations on the marketplace. However, it can also be enforced through technology, but it depends on the architecture of the marketplace. *E1* gives an example where an ***authority*** facilitates the ***contract*** of data exchange among data actors. *E1* says that contracts define the business process of using technology to carry out *data exchange*. So, *E1* says the requirement of *secure data exchange* is also governed by the governing authority, stressing that an authority actor is necessary for ***governance,*** and that *secure data exchange* is a part of *governance* requirement. On the other hand, *E2* reflects on our description of data governance and states that "it is a combination of the *secure data exchange, data sovereignty* and *data provenance*". This statement relates to what *E1* stated earlier that *governance* involves managing all the activities of the data marketplace.
- ***Data Economy:*** *E2* agreed with our description of *data economy* saying that it is fine to be a requirement. *E1* did not touch on this.
- ***Data Sovereignty:*** *E1* thinks that it is true that data sovereignty can be enforced but it depends on the design. He says, in a *centralized* design, the central authority has the control over data and sovereignty here means that the data owner trusts the central authority to do what the owners asks him to do. But it can be truly enforced in *decentralized* design by keeping the data on blockchain where the data owner can control it. However, if the data is copied, then *data sovereignty* is lost. But since there is no real life blockchain application on this yet, *E1* says this is a direction to investigate. *E2* thinks of *data sovereignty* as a requirement to be fine. However, *E2* reflects again on our phrasing in the description of *data sovereignty* that it is not about protecting the *legality of the data* as "*the data is either legal or illegal*". *E2* suggests that *data sovereignty* is basically having ***control*** over who uses the data.
- ***Secure Data Exchange:*** *E1 and E2 were fine with our description of the secure data exchange and it being a requirement.* However, *E1* had a concern relating to this subject that "*once the consumer gets the data, nothing stops him from doing whatever he wants with the data*". *E1* says this issue as the more pressing issue than an external entity intercepting the transacted data. *E2* had a phrasing issue over the consistency of the term ***data actors*** as we used inconsistently with the terms "*data customer*", "*data subjects*".
- ***Data Exchange Platform:*** *E2* found this requirement to be redundant as it is the complementary requirement of rest of the requirements.

Reflecting on overall of requirements, *E1* remarked that ***governance*** is the fundamental requirement and the rest of the requirements is dictated by the use-case and the architecture of the data marketplace platform. On the other hand, *E2* reflected that the requirements are "*reasonable*" to have for a data marketplace platform; while also suggested that these requirements are "*exhaustive in the sense that they are generic*" and the requirements cover all the bases relevant to a data marketplace.

6.2.2.2 Conclusions

Although the theoretical concepts associated with our requirements were only from technological standpoint, we realized we should include the non-technological aspects to have a comprehensive understanding of the requirements. This was also recommended by expert, *E1* as he said it is not possible just with technology alone, but we need a non-technological governing authority to effectively achieve the fundamental functioning of the data marketplaces. Furthermore, the interpretation of each

requirement was also validated and are refined here as applicable to reflect the credible expert insights. Furthermore, after the analysis, it was deduced that the functional requirements should provide objective description of the requirements applicable to the data marketplace platforms. As a result, when describing the functional requirements here, we omitted from the description, the examples of how they are enforced by the data marketplaces as they are implementation-dependent but not objective information.

- ***Boundary Conditions:*** The description of the boundary conditions remains the same as before which is, "*Strict boundary conditions help in authorizing only the legitimate costumers willing to share or buy data. This helps in safeguarding the data from unauthorized access*".
- ***Data Provenance***: This requirement undergoes changes in its description where we omit the phrase "*change of ownership*" as the concept was disregarded by *E2.* Although, considered as a possibility by *E1,* it is never observed to be in practice. The concept that does exist in data marketplace is the concept of *licenses*. Practically, the data owner always owns the data and, he provisions the data to the data consumer who can use it according to the terms agreed in the licensing contract. So, we change the phrase "*change of ownership*" to "*data usage*" which is actually in lines with the meaning of data provenance. So, the description changes to "*Data Provenance is a requirement to track and document the **data lineage** and **data usage**. Data lineage refers to the transformation of the data from its original state to the current state (different versions). Data usage is focused on who has the access to the data, who accessed it and if they accessed or not".* The metadata aspect is omitted from the description here as the enforcement of data provenance is implementation dependent and is more a part of functional components which deals with features like that of metadata.
- ***Data Economy:*** This requirement remains the same too which "*reflects the business purpose of the data marketplace platform which is to generate revenue stream for itself through its services"*. However, we have excluded the information about its way of implementing.
- ***Data Sovereignty:*** After discarding the phrasing of "*legality of data*", this requirement can be described as *a mechanism expected for the data marketplace platform to support for the data owner to have control over his data and its usage".* We omit implementation examples.
- ***Secure Data Exchange:*** There is no change in interpretation of this requirement. Its description remains the same as "*the most fundamental aspect of the data marketplace platform which is carry out the data exchange between the data actors in the most secure way".*

The requirement of *Data Exchange Platform* is removed from the list as it is declared as redundant. Moving on, the new requirements that evolved from the expert insights and further analysis were also included into the list. These are described as follows,

- ***Marketplace Platform***: This requirement is a transformed version of the *Data Exchange Platform* which basically deals with the platform aspects of the data marketplaces which is obviously a fundamental requirement for a data marketplace platform. This requirement is described as "*the requirement of platform features like match-making between the participants; and the marketplace features like cataloguing, curation, e-commerce mechanism, recommendations et cetera*". This description makes way more sense than the previous one of *Data Exchange Platform* and hence, also makes it different.
- ***Legal Management:*** This requirement is for the data marketplace platform to handle the legal aspects of data trading like *contract management, license management, litigation* etc. This requirement is enforced by a human actor and not by technology.
- ***Trust Mechanism:*** This is also a non-technological requirement enforced by a different kind of human actor which is more like an independent authority, for example, *Certification Authority, Auditing Authority* et cetera; who through their operations, create trust for the data actors to participate in data trading over the data marketplace platform.

The requirements of Legal Management and Trust Mechanism cannot necessarily be enforced by the data marketplace platform itself but can be done on an ad-hoc basis by external entities which possess expertise of specific issues like *Certification, Auditing, Legal Counsel* et cetera. Furthermore, *Legal Management* and *Trust Mechanism* can currently be enforced purely by authority actors on the data marketplace platform; while the rest depend on their implementation consisting of a coordinated effort of both technology and actors. However, cutting-edge technologies like Blockchain, Multi-Party Computation (MPC), Homomorphic Encryption et cetera can enable the data trading technologically

alone without any human actor. But this is just a claim as the said-technologies have not achieved the desired level of sophistication to be applied in real-life cases. Evidently, investigating this claim is part of our research problem but we are only doing it with respect to MPC technology.
Moving on, the above list represents the updated functional requirements and we categorize all of these under a core category reflecting the most fundamental requirement for a data marketplace platform to satisfy which is, ***Governance***. *Governance* can be described as the requirement of a mechanism which oversees all the activities on the data marketplace. As specified by *E1*, it can only be enforced by the right coordination between the human actor and the technology; however, difficult with one of them alone. Furthermore, subcategories were created for this core category. Since the requirements of *Data Provenance*, *Data Economy*, *Data Sovereignty* and *Secure Data Exchange* relate to the overseeing of the activities associated with data, we group these requirements under the subcategory, ***Data Governance.*** On the other hand, we group *Boundary Conditions*, *Marketplace Platform*, *Legal Management* and *Trust Mechanism* under the category of ***Marketplace Governance*** as they comprise of overseeing the activities specifically of the marketplace aspect. The updated list of categories and codes reflecting the refined functional requirements is listed below and is illustrated in **Figure 15** in the form a hierarchy under the core category of *Governance*.

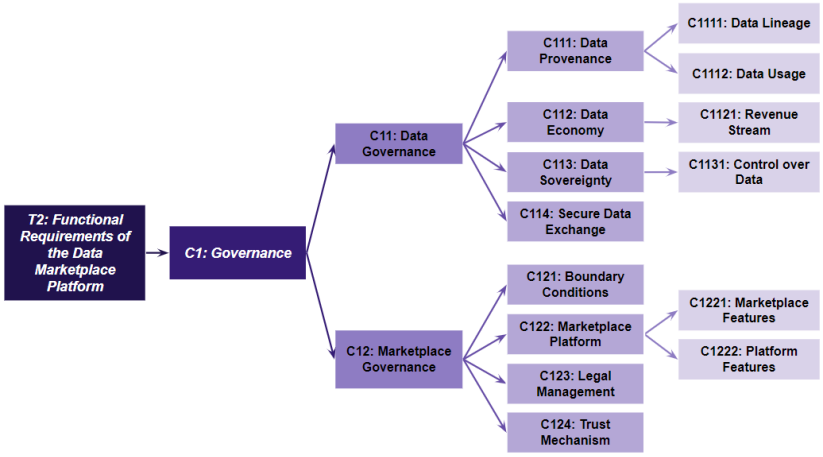

Figure 15: Functional Requirements 2.0 of the Data marketplace Platform

6.2.3 Customers of the Data Marketplace Platform

The list of customers dealt in Chapter 2 was not an exhaustive list. Hence, the topic, *T3: Customers of the Data Marketplace Platform* was included as an exploratory study so that we can validated and updated the list. The theoretical concepts associated with this topic were analyzed by relating them to the insights of experts *E1* and *E2*. The initial list of categories and codes in this topic derived from Chapter 2 were:

- ***T3:*** *Customers of the Data Marketplace Platform*
- ***C1:*** *Data Providers*
  - ***C11:*** *Data Collectors*
  - ***C12:*** *Data Managers*
  - ***C13:*** *Data Aggregators*
- ***C2:*** *Data Consumers*

6.2.3.1 Results & Analysis

In terms of the actors, *E1* stresses the significance of an ***authority*** who is according to him, very crucial for the governance of the data marketplace. *E1* also mentions different *authorities* which carry out different functions in the data marketplace like *Certification Authority, Auditing Authority* et cetera. *E2*

suggests on maintaining the consistency of the terminology in the descriptions with what is used in the industry like ***data actors*** instead of *data subjects*.

6.2.3.2 Conclusions

With the refinement of our focal data marketplace to *many-to-many B2B serendipity model*, the updated list of functional requirements combined with the expert insights, we decided to include further actors into the architecture who are not customers but play a crucial role for the functioning of the data marketplaces. As a result, we renamed the core category from "*Customers*" to ***Actors of the Data Marketplace Ecosystem*** to reflect the ecosystem view of the data marketplaces.
To maintain the terminology consistent with the industry usage, we modified the core categories of Data Providers and Data Consumers into a single category named as ***Data Actors*** which reflect the customer definition from the initial list. Further sub-categories were added; namely, ***Data Supply*** side and ***Data Demand*** side which are consistent with the industry usage. In the *Supply side*, we included the actors who supply data and data related services on the data marketplace platform; basically, *Data Owners, Data Managers, Data Aggregators* and even third-party data analysis service providers. On the *Demand side*, we put *Data Consumers.* All the actors retain their previous interpretations from the initial list in the sense that they use the platform services for their benefit.
Apart from these *data actors*, we also included the actors who enable the data marketplaces like the *authority services* as stressed by the experts. We termed these actors as ***Marketplace Enabling Actors***. These actors represent the ***network*** aspects where the criteria for the actors expands from the usage and non-usage of the services to the creation and capture of ***value*** in the data marketplace system. We further divided the enable actors into 2 categories:

- ***Marketplace Provider:*** This actor is the central authority who provides the data marketplace service by hosting and managing all the services and operations on a technological platform. This is an organization whose business model is to provide the data marketplace service and enforces the requirement of *Governance* by implementing the business processes using either technology or just human actors.
- ***Independent Service Providers:*** These actors are independent actors who provide services to enable the data marketplaces as and when necessary. The services can range from technological services like infrastructure provider to non-technological services like certification, auditing, legal counsel et cetera. Mostly, these actors enforce the non-technological requirements like *Legal Management, Trust Mechanism* et cetera.

The updated list of the categories and codes reflecting the actors in the data marketplace ecosystem is illustrated in **Figure 16**.

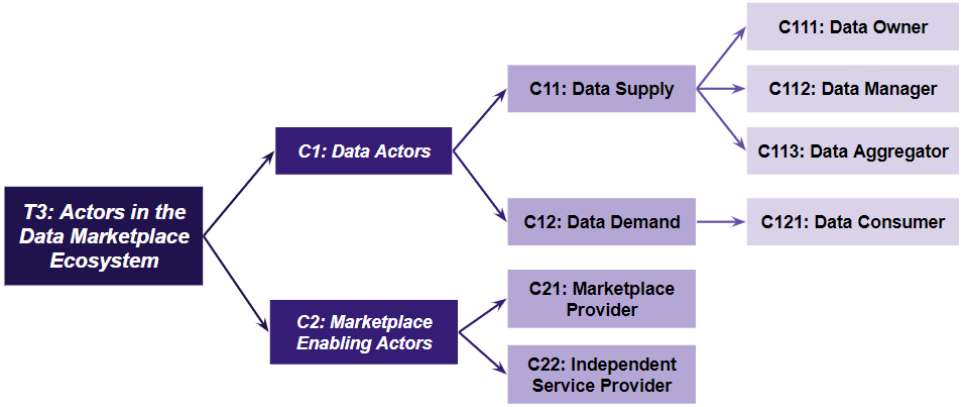

Figure 16: Actors 2.0 in the Data Marketplace Ecosystem

6.2.4 Functional Components of the Data Marketplace Platform

The functional components developed in Chapter 2 were the result of our desk research and hence, they were needed to be validated with expert insights to refine into more valid components. Hence, we decided to include topic *T4: Functional Components of the Data Marketplace Platform* for an

exploratory study where the related theoretical concepts can be analyzed by relating them to the experts, *E1* and *E2*. The initial list of categories and codes were:

- ***T4:*** *Functional Components of the Data Marketplace Platform*
- ***C1:*** *Identity Management*
  - ***C11:*** *Features*
    - ***C111:*** *Induction*
    - ***C112:*** *Authentication*
    - ***C113:*** *Authorization*
  - ***C12:*** *Enforced Requirements*
    - ***C121:*** *Boundary Conditions*
- ***C2:*** *Broker Service*
  - ***C21:*** *Features*
    - ***C211:*** *Data Management Services*
      - ***C2111:*** *Data Cataloguing*
      - ***C2112:*** *Data Marketplace Curation*
        - ***C21121:*** *Data Categorization*
        - ***C21122:*** *Data Tagging*
      - ***C2113:*** *Data Tracking*
        - ***C21131:*** *Data Lineage Tracking*
        - ***C21132:*** *Data Usage Tracking*
    - ***C212:*** *User Interaction Services*
  - ***C22:*** *Enforced Requirements*
    - ***C221:*** *Data Exchange Platform*
    - ***C222:*** *Data Governance*
    - ***C223:*** *Data Provenance*
    - ***C224:*** *Data Economy*
- ***C3:*** *Clearing House*
  - ***C31:*** *Features*
    - ***C311:*** *Transaction Repository*
  - ***C32:*** *Enforced Requirements*
    - ***C321:*** *Data Provenance*
- ***C4:*** *Data Inventory*
  - ***C41:*** *Features*
    - ***C411:*** *(Meta)Data Storage*
  - ***C42:*** *Enforced Requirements*
    - ***C421:*** *Data Governance*
    - ***C422:*** *Data Sovereignty*
- ***C5:*** *Data Exchange Service*
  - ***C51:*** *Features*
    - ***C511:*** *Data Exchange Mechanism*
  - ***C52:*** *Enforced Requirements*
    - ***C521:*** *Secure Data Exchange*
- ***C6:*** *Data Analytics Service.*
  - ***C61:*** *Features*
    - ***C611:*** *Data Analytics Tools*
  - ***C62:*** *Enforced Requirements*
    - ***C621:*** *Data Economy*

6.2.4.1 Results & Analysis

The experts reflected on our conceptualized components one by one and their comments on each component is discussed below.

- ***Identity Management*****:** *E1* suggests that if we are dealing with a decentralized data marketplace, then data exchange happens through an encrypted channel involving public key encryption. Since the data marketplace is responsible for the data exchange, it should generally have a ***key***

management system to manage keys and in turn, the communication channels. Apart from this, both the experts *E1* and *E2* were "*fine*" with our features of *Induction, Authentication* and *Authorization.*

- ***Broker Service:*** *E1* validates that we "*need catalogues of data objects and metadata of each object to describe the data that is being showcased on the data marketplace platform*". Furthermore, *E1* says the management of *physical data* is also done by the *broker*. *E2* says that in Data Market Austria, they separate *metadata* and the *data*; the *metadata* is centralized and is completely relied on by the *broker service*. *E2* further reflects on the feature of *data tracking* and reflects that only *data lineage* can be part of the *data tracking* feature, while the *data usage* is more applicable in the context of ***transaction management***. Other than that, both experts were okay with the rest of features of the *broker service.*
- ***Clearing House:*** *E1* stressed that since the *transactions* are needed for auditing purpose, the management of *transactions* is crucial for the data marketplace. Hence, *E1* suggested that this should be implemented using some tamperproof database or blockchain ledger. *E2* reflected that the feature of *data usage tracking* should be integral to the *clearing house* component.
- ***Data Inventory:*** *E1* shared his skepticism on how the decentralized design can be materialized. *E1* explained a possibility with the help of *key management system*. *E1* said that *data can be on the provider's site* and if we want to compute something on the data, we can have a container with an algorithm which needs to decrypt data. So, *E1* said it goes to *key management* again to manage the credentials of the data. *E1* also stated that however, the container with the algorithm can copy the data for itself which breaks the security. So, *E1* said we need ***governance model*** to manage this situation. On the other hand, *E2* pointed out the *data management* feature is part of *broker service* and it does not make sense to have a *data inventory* component. Hence, *E2* suggested that we can exclude *data inventory* component. *E2* also remarked that it does not matter where the physical data resides as it can be stored on a distributed system and its access can be managed by *broker service.*
- ***Data Exchange Service*****:** *E1* did not have any comments here except for perceiving it just as a ***communication channel*** enabling *secure data exchange*. However, *E1* reflected on different aspects of designing the business process of the data sharing among the actors; some of which were: how the infrastructure of the data sharing is designed, whether the parties have preferences there, how the data access is provided, whether through algorithm or a container. *E1* suggested all these aspects to be related to *secure data exchange* and hence, can be part of this component. *E2* expressed his problem with this component as he understood that significant processes involved in data trading have been taken care of by the previous components. In that light, *E2* states that mentioning this service just as a "*download link with SSH*" as a very basic thing to explicitly describe. *E2* remarked that if by data exchange service is interpreted as the network, a connection between two end points like saying, "*internet is part of the data market*" which is a very trivial thing in this discussion.
- ***Data Analytics Service:*** *E1* perceived this as ***data analysis service*** being hosted on the *data marketplace platform* or ***a third-party cloud provider***. In that case, *E1* suggested to have a credential management to verify the legitimacy of these entities before providing access to carry out data analysis on the data. *E2* remarked data analysis services part as very important and suggests 2 variants of provisioning the data analysis services: one variant where *data analysis services are* ***centralized*** and run on the platform and the other variant being the one with *third parties* offering *data analysis services* in an ***app store*** kind of way. Consequently, *E2* was okay with our app store model. However, we change the name of this component to ***Data Analysis Service*** to reflect both interpretations.

6.2.4.2 Conclusions

As the focal data marketplace platform of this research was specified to be *many-to-many B2B Decentralized Serendipity Model data marketplace*, the centralized platform design was excluded from the analysis thus narrowing our scope. As a result, the usage of the term (*meta*)*data* to signify both data and metadata being on the platform is no longer used. Furthermore, only *metadata* is managed on the

platform centrally while the data resides decentralized. Below, we provide the updated list of functional requirements:

- ***Identity Management***: The interpretation of this component remains the same with the features of *induction, authentication and authorization.* However, a new feature is added i.e. ***key management*** as this is a crucial requirement for the materialization of decentralized data marketplace platform for the enabling and management of encrypted communication channel. Evidently, this component enforces the functional requirement of only *boundary conditions.*
- ***Broker Service:*** This component contains the same 2 features: *Data Management* and *Customer Interaction*. Some of the activities which are part of data management feature remain same while some undergo changes. *Data Cataloguing* and *Data Marketplace Curation* remain the same. *Data Tracking* undergoes a small change with only handling the tracking of *data lineage*. Hence, we rename it as ***Data Lineage Tracking***. Finally, since data does not reside on the platform, the broker service is responsible only for the *management of providing access to the appropriate data* wherever it resides (either on data owner's site or in a distributed system) to the appropriate actors with the help of ***key management***. We term this activity as ***Data Access Management.*** On the other hand, there are no changes in the user interaction service. The *broker service* enforces the following functional requirements
  - *Data Provenance* through *data lineage tracking*;
  - *Data Economy* by creating revenue streams for themselves and the actors.
  - *Marketplace Features* though their *data management services*
  - *Platform services* through *user interaction service.*
- ***Clearing House:*** The interpretation of this component also does not undergo any change as it essentially comprises of transaction management system. The component enables *data usage tracking* which involves documenting the usage information of the data like *who has the access to the data, who accessed it and if they accessed or not et cetera.* With this activity, clearing house enforces *data provenance* functional requirement. It can be implemented in different ways. Although the underlying condition is that it should be tamperproof.
- ***Data Exchange Service:*** This component undergoes a major change as a result of the expert insights as our conceptualization of this component was unclear and very trivial to be a functional component. This component is no longer just a communication channel or a download link with SSH. The data exchange service signifies the ***business process of how the data is shared*** among the involved data actors. In simple words, the logistical way through which the ***data access*** is provided to the data consumer on the data marketplace. The implementation of this component is highly dependent on the use-case and resulting technical architecture. The concepts like computation, algorithm, data access and even data analysis comes into the picture based on the underlying use-case of data-sharing. With these aspects, the data exchange service enables the functional requirement, *secure data exchange.* Additionally, it goes without saying that in a decentralized design like ours, the data owner has control over his data as he houses the data and access is provided by the data marketplace to the data consumer which is dictated by the use-case of the data sharing. Since this aspect relates to the mechanism of data sharing, this component also enforces the functional requirement of *data sovereignty.*
- ***Data Analysis Service:*** We shall incorporate the additional insight on this component which we got from the experts that *data analysis services* can be also be hosted ***centrally*** on the data marketplace platform. Again, the way to do it is dependent on the business process of the data analysis which is dependent on the use case and the architecture. The feature of app store model still remains with the platform providing data analytic tools from the third parties on the platform in the form of downloadable software or SaaS. The functional requirement of ***data economy*** is satisfied here.

The ***Data Inventory*** component is omitted from our list for a variety of reasons. Firstly, the platform design is decentralized and hence the data does not reside on the data marketplace platform; consequently, eliminating the need for data inventory. Furthermore, in a decentralized setting, where the physical data resides, whether on the client's site or a distributed system or in rare cases in blockchain, does not matter from the perspective of the data marketplace as it is the responsibility of the data owner

provisioning the data. The owner provides the access of the data to the broker service which manages that access. These reasons motivated us to remove the component from the list.
In addition to the existing components, we included a new component, ***Governance Model*** to the list of functional components. As discussed earlier in the requirements sections, this component consists of activities which involve enabling the data marketplace platform in the form of trust mechanism, governance or enabling services. These activities are carried out by the *Market Enabling Actors* by designing business processes using technology. Consequently, it fulfils the functional requirements of ***Governance.*** The enabling services can be added as and when necessary according the use-case. Hence, the actors and activities here are not fixed. The following is the updated list of categories and codes associated with the topic, *T4*: *Functional Components of the Data Marketplace Platform* in which the modifications highlighted (additions in ***green*** and deletions in ***red***):

- ***T4:*** *Functional Components of the Data Marketplace Platform*
- ***C1: Governance Model***
- ***C1:*** *Identity Management*
  - ***C11:*** *Features*
    - ***C111:*** *Induction*
    - ***C112:*** *Authentication*
    - ***C113:*** *Authorization*
    - ***C114:*** *Key Management*
  - ***C12:*** *Enforced Requirements*
    - ***C121:*** *Boundary Conditions*
- ***C2:*** *Broker Service*
  - ***C21:*** *Features*
    - ***C211:*** *Data Management Services*
      - ***C2111:*** *Data Cataloguing*
      - ***C2112:*** *Data Marketplace Curation*
        - ***C21121:*** *Data Categorization*
        - ***C21122:*** *Data Tagging*
      - ***C2113:*** *Data Lineage Tracking*
      - ***C2114:*** *Data Access Management*
    - ***C212:*** *User Interaction Services*
  - ***C22:*** *Enforced Requirements*
    - ***C221:*** *Data Provenance*
      - ***C2211:*** *Data Lineage*
    - ***C222:*** *Data Economy*
      - ***C2221:*** *Revenue Stream*
    - ***C223:*** *Marketplace Platform*
      - ***C2231:*** *Marketplace Features*
      - ***C2232:*** *Platform Features*
- ***C3:*** *Clearing House*
  - ***C31:*** *Features*
    - ***C311:*** *Transaction Repository*
    - ***C312:*** *Data Usage Tracking*
  - ***C32:*** *Enforced Requirements*
    - ***C321:*** *Data Provenance*
      - ***C3211:*** *Data Usage*
- ***C4:*** *Data Exchange Service*
  - ***C41:*** *Features*
    - ***C411:*** *Data Exchange Business Process*
  - ***C42:*** *Enforced Requirements*
    - ***C421:*** *Secure Data Exchange*
    - ***C422:*** *Data Sovereignty*
- ***C5:*** *Data Analysis Service.*
  - ***C51:*** *Features*

        - ***C511:*** *Data Analysis*
        - ***C512:*** *Data Analytics App Store*
    - ***C52:*** *Enforced Requirements*
        - ***C521:*** *Data Economy*
- ***C6:*** *Data Inventory*

Following the updating of the three components of the high-level architecture of the data marketplace from Chapter 2, a new updated high-level architecture is built to reflect the findings obtained so far and represents a more appropriate and comprehensive architecture for a data marketplace platform. The updated architecture is illustrated in **Figure 17** which represents the ***Pre-MPC Data Marketplace Platform 2.0***.

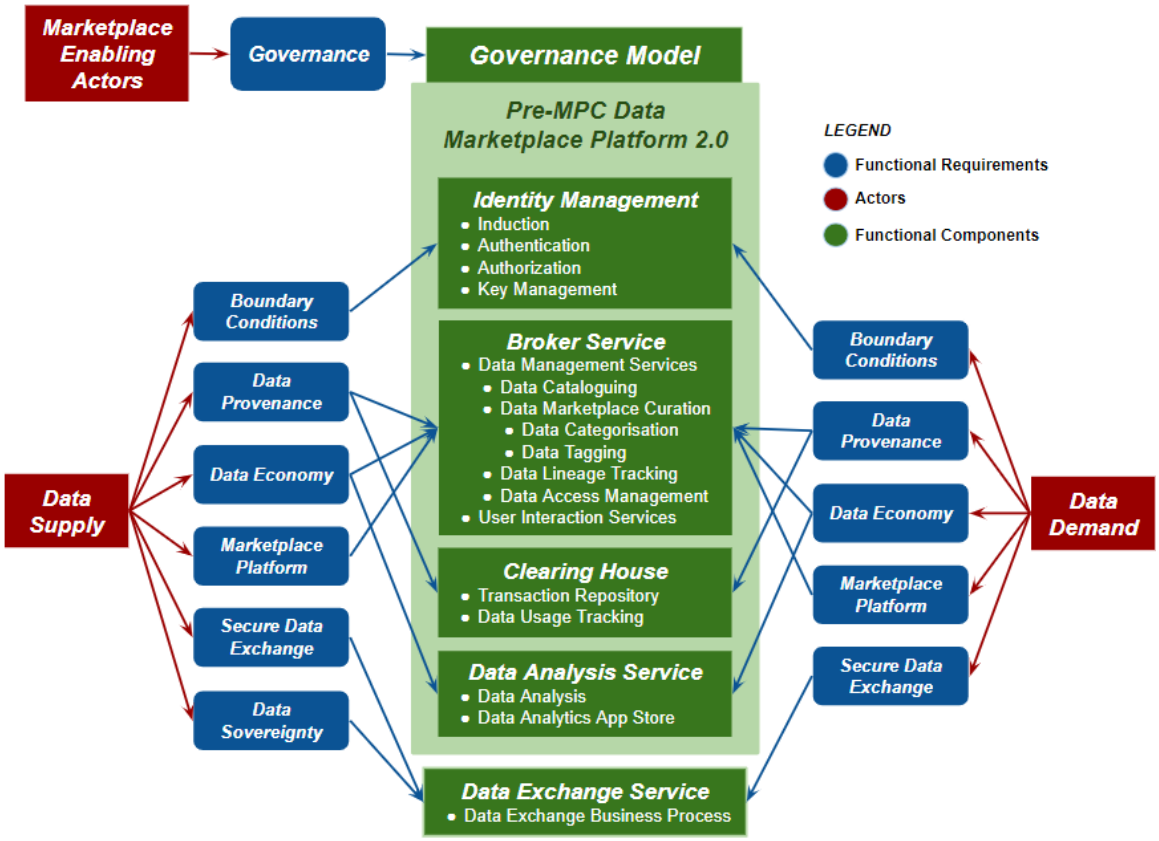

Figure 17: Refined High-Level Architecture of the Data Marketplace Platform (Pre-MPC Data Marketplace Platform 2.0)

Governance Model and *the Data Exchange Service* are intentionally placed outside the platform in Figure 20. The *governance model* comprises of human actors and activities which enforce *governance* on the data marketplace platform by devising various business processes using technology. So, the governance model reflects the coordination between the human actor and technology which collectively enable the functioning of the data marketplace platform. Hence, it did not make sense to include governance model inside the technological architecture of the data marketplace platform. On the other hand, the *data exchange service* is an ad-hoc component which is materialized outside the platform between the data actors involved in the use-case relationship which was established over the platform.

6.2.5 *HLA* Framework

The theoretical concepts of *T6: HLA Framework* from Chapter 2 were not explicitly considered for validation during the expert interviews. However, the updating of the high-level architecture of the data marketplace platform brought about significant changes in the functional requirements, actors and

functional components. Hence, it was decided to translate these changes to update the specification of the attributes to obtain an updated HLA framework. The initial list of codes derived from Chapter 2 were:

- ***T5:*** *HLA Framework*
- ***C1:*** *Functional Requirements*
- ***C2:*** *Customers*
- ***C3:*** *Functional Components*

The overall change that the architecture underwent was with respect to its scope. It was understood from expert insights of *E1* that the operations of any technological entity like data marketplace platform cannot be materialized technologically alone but needs a coordinated marriage between human actors and technology. Hence, the scope of the architecture was expanded not only to include the focal technological entity but also the ecosystem that enables the technological entity; basically, the human factor associated with the enabling of the focal technological entity.
This change in scope can be propagated to HLA framework as the ecosystem view of the technological entity is more insightful for analysis than the technological one alone. Essentially, the resulting high-level architecture of a technological entity obtained from *HLA framework* will reflect the ecosystem (comprising of human factor) in which the focal entity operates along with its technological architecture. This change brought about changes in all the attributes which reflect the increased scope.

- ***Functional Requirements:*** The modified interpretation of the *Functional Requirements* now reflect not only the technological requirements but also technological ecosystem requirements which reflect the expectations of the actors in the ecosystem from the focal technological entity.
- ***Actors in the Ecosystem:*** The previously termed, *Customer* attribute undergoes major change to expand the horizon to include the human actors along with the customers who enable the focal technological entity. Hence, the attribute was renamed into *actors in the ecosystem*.
- ***Functional Components:*** Similarly, the components comprising of the human activities like auditing, trust enforcement et cetera are also included here now which could give rise to components comprising either solely the human activities or an amalgamation of human and technological activities.

However, the definition of the result architecture would not undergo any change as the it still provides an architecture to a technological entity with surface-level information but not technical specification which applies for either of the technological and human activities. The modified framework is illustrated in **Figure 18**.

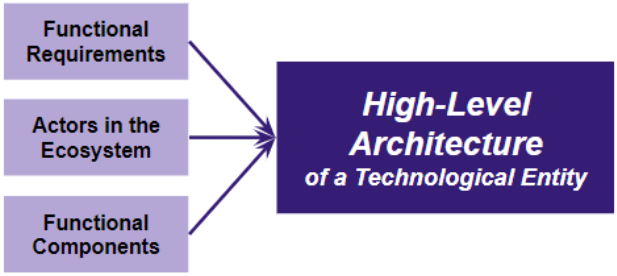

Figure 18: HLA Framework 2.0

6.3 Validation of Post-MPC Data Marketplace Platform 1.0

The topics and corresponding theoretical concepts of the research focus, *RF2* are validated here. The artefact under consideration here is the *Post-Data Marketplace Platform 1.0*. The following two topics are validated in the upcoming subsections.

- ***T6***: *Perception of MPC Technology*
- ***T7:*** *MPC Incorporation into the Data Marketplace Platform*

6.3.1 Perception of MPC Technology

Since we are not experts in the technical aspects of MPC, our perception of MPC technology is based only on how Safe-DEED describes it in their project proposal and the same was understood and incorporated into our study. For this reason, this topic was included here so that our conceptual perception could be validated from the experts and thereby, make the further analysis valid. The validation activity was carried out with the insights obtained primarily from the expert, *E4* who is specialized in MPC technology and works in Safe-DEED to implement the technology. *E4* dealt extensively with the value of MPC for data economic market in general. Additionally, experts *E1* and *E2* also provided their insights in this subject which reinforced the insights of *E4*.

6.3.1.1 Results & Analysis

When asked to describe what Multi-Party Computation is, *E4* explains that the basic idea is to bring different parties together to compute something on their inputs without the parties knowing about the inputs of rest of the parties; ultimately learning only the result of the computation and nothing else. But *E4* says that generally this happens with a trusted authority who takes the inputs, computes the function and gives back the result of the computation. Consequently, the authority learns the input data from all the parties. *E4* says that MPC can transforms process into a protocol where the protocol executes the computation, essentially eliminating the trusted authority and still getting the same security guarantee that the result is computed and sent to a dedicated party; without the parties knowing the inputs of the rest of the parties.

E4 relates to the advantages of this property of MPC by mentioning the following. Firstly, the concept of trust is enforced by the system itself and not the actor as there is risk involved. Secondly, *E4* says that with MPC, we can work on data without having to worry about "*leaking personally identifiable information"* in the process. Consequently, *E4* says, "*we wouldn't even need any anonymization techniques because you don't actually have to send the data*" and it is shared through a protocol "*in a randomized way that the others can't learn anything from it*". Adding to this, *E4* further suggests that using MPC, we can carry out computations on private data that is sensitive and that is not legally possible to combine with other data like the "*data from health insurance companies with hospitals as they can't share their databases*". *E4* remarks that the rules around these databases restrict the involved actors to just send their databases to other parties to combine them and compute statistics like "*how often is a person sick? Or are there any other trends like people with higher education get sick less often"*. However, *E4* suggests that MPC allows to compute these statistics because "*the data never leaves your premises in a way that the other party can decrypt it" but* is given access to a protocol that runs the computations and gets only the result. In addition to this, *E4* provides further examples of interesting applications where the property of MPC comes in handy which include "*an auction system where the bids stay private until the final bid is decided*".

Moving on to the logistics of designing and implementing a business process with MPC technology, *E4* states that it starts with a use-case where it makes sense for the companies to interact and share data for which MPC can enforce security for leakage of sensitive private data or confidential proprietary data that is internal to the companies. *E4* provides an example of a use-case where two companies can combine their customers lists to generate products interesting for the customers in common. Since the list of customers is a confidential proprietary information, they cannot be combined in a traditional way but MPC enables this with one of its protocol called private set intersection (PSI). *E4* stresses that use-case is critical to have beforehand as it will direct the decisions like choosing the protocol, designing the process and running the protocol.

When asked about the two MPC processes conceptualized in Chapter 5, for the interactive process, *E4* confirms that it is a valid process but basic one as different variations of this is possible where everybody receives the output or somebody that is not involved receives the output or some actor only providing computation service over cloud but not providing any data. About the non-interactive process, *E4* disregards the process to be of MPC but rather of traditional computation involving another privacy preserving technology, homomorphic encryption. *E4* reflects that the non-interactive one is a valid process of data aggregation which enables the data owner to provide his data once and not be present every time the computation happens. However, since it is not of MPC technology, the process is out of

our scope. Reflecting generally of the processes, *E4* suggests that homomorphic encryption can also be part of the MPC protocol; even data analysis can also be defined as part of the MPC protocol. However, the underlying use-case decides whether the former should be part of the protocol. On this subject, even *E2* reflected confirming that the interactive process is valid representing the true promise of MPC and states that there are many different models of processes which are being developed by his colleagues at *Safe-DEED*.
Coming to the limitations of MPC technology, *E4* reflects that the MPC protocol is driven by the function from the use-case. So, it should be made sure "*the [MPC] function needs to have the property that if you have the input and the function output, then you don't learn anything about the other inputs";* basically, reverse engineering of input data should not be possible from outputs in conjunction with the MPC function. Related to this topic, *E1* also remarks that the application of MPC is currently limited in the real world, due to implementation concerns.

6.3.1.2 Conclusions

The insights about the basic idea, properties and the advantages of the MPC technology were consistent with what we had dealt. However, the discrepancy with the perception arose in case of processes defined in Chapter 5. It was presumed that the two processes represented two kinds of processes of MPC. But it turned out that only interactive process was of MPC and non-interactive was not. However, the valuable insight gained in this topic was that of MPC protocol being designed based on an underlying use-case. The use-case being that of data sharing among companies which were suggested earlier by the experts. The fact that the underlying use-case of data computation directs the selection of the function and the design of MPC protocol clarifies that the MPC technology is designed in an ad-hoc way as required by the use-case. This falsifies our perception that MPC is a fixed process like the two processes mentioned in Chapter 5 and that they must be used that way by the actors. On the contrary, the protocol is designed as required by the use-case of the actors. Another useful insight is that the protocol can contain other constituents like homomorphic encryption, different kind of data analysis functions etc. Hence, MPC can carry out many functionalities of data in addition to enable data sharing in a confidentiality-preserving and privacy-preserving way.
However, MPC technology has its own limitations. Firstly, it is still in conceptualization phase and has not reached maturity as it suffers from scalability issues. Another limitation is that, it is unknown if every function or computation is compatible to be converted into an MPC protocol. The functions derived out of the underlying use-case should be compatible with ***Safe-DEED Primitives*** to be converted into a valid protocol. All these limitations should be explored in the future to bring the promises and potential of MPC technology to reality.

6.3.2 MPC Incorporation into the Data Marketplace Platform

This topic represents the first of the two conceptual models contributing towards our research objective as validation of this topic contributes towards the understanding of the architectural implication of MPC technology to the data marketplace platforms. The concepts associated with this topic were analyzed by relating them to the insights predominantly of experts *E4* but also, *E1* and *E2*. The contents of section 5.2 drove the list of categories and codes which are listed below.

- ***T7:*** *MPC Incorporation into the Data Marketplace Platform*
- ***C1:*** *Powers Data Exchange Service*
  - ***C11:*** *Safe-DEED Component*
    - ***C111:*** *Safe-DEED Primitives*
    - ***C112:*** *Safe-DEED Network*
- ***C2:*** *Enables Decentralised Design*
  - ***C21:*** *Changes Data Inventory to Metadata Inventory*
  - ***C22:*** *Moves Data Sovereignty towards Data Provider's site*
  - ***C23:*** *Moves Data Governance to Data Provider's site*

We first discuss how MPC technology can be applied generally in data marketplace and then later, validate the incorporation for the updated architecture of the data marketplace platform, ***Pre-MPC Data Marketplace Platform 2.0***.

6.3.2.1 Results & Analysis

When asked about what the application of MPC in a data marketplace is, *E4* remarked that the idea of MPC that can work in data marketplaces is that data marketplace can be a platform, where the data owners can say that they have some data and the parties interested in using or running some analysis on that data can connect with the data owner; and then they both can run the MPC protocol privately between them. Evidently, *E4* says that data marketplace can be a place where companies find each other and establish relationship, and the connected companies can install ***Safe-DEED Component*** containing the MPC protocol on either of their servers and can carry out data computation. On this subject, *E1* remarks that with MPC, the system itself provides security where the data owners have full control over their data and thereby reducing the need for security governance. *E1* specifically says to enforce decentralized design, MPC makes a huge difference as it eliminates the need for *key management* and the risks associated with it. Sharing this thought, *E2* also says that MPC will play a role in enforcing data sovereignty as the data can no more be misused by the data consumer.
Regarding the changes that MPC technology can bring about in our architecture, *E4* reflects that the components which undergo change with the incorporation of MPC technology would be: ***Data Exchange Service*** and ***Data Analysis Services***. *E4* continued that data exchange service will be transformed with MPC Technology which is enabled by ***Safe-DEED Component*** de-centrally running on the connected parties' servers. On the other hand, data analysis service will be moved to the sites of the parties (data owners, data aggregators and data consumers); away from the platform as the data analysis services are run as part of the MPC protocol itself. Other than that, *E4* states that MPC would not affect any other component. *E2* suggests that the data exchange service will be transformed into safer than the traditional way, while also reflecting that none of the other components undergo any change.

6.3.2.2 Conclusions

Here, we shall reflect what the above findings mean to our research and incorporate the appropriate changes in the updated high-level architecture of our data marketplace platform. The foremost conclusion on the application of MPC technology (foregoing its limitations) is that it enables a truly decentralized data marketplace platform by truly enabling data sovereignty for the data owners. Furthermore, MPC technology provides *security-by-design* as propositioned in Chapter 5 by truly enabling data sharing and data analysis services in a confidentiality-preserving and privacy-preserving way (where actual data is not known to anybody other than the one who owns it).
The changes brought about with the incorporation of MPC technology into the updated *high-level architecture* are as follows:

- The ***Data Exchange Service*** gets transformed from a traditional process (SSH encrypted channel) to a safer and more sophisticated process by including MPC technology through ***Safe-DEED component*** (***Safe-DEED Primitives*** and ***Safe-DEED Network***). The data exchange service will be designed in an ad-hoc way which will implemented in the form of an MPC protocol executing the computation through the ***Safe-DEED Component*** running on the servers of all the involved parties.
- The ***Data Analysis Service*** becomes a feature of Data Exchange Service as the data analysis becomes part of the MPC protocol. However, the ***App Store*** component remains on the platform which provides data analytics tools to the actors in the form of downloadable software or SaaS model. So, we shall rename this component as ***Data Analytics AppStore*** to signify its actual meaning.
- The *key management system* in the Identity Management remains but its involvement in the data exchange service depends on MPC protocol if it contains encryption elements.
- Finally, the responsibilities of the security aspect and the trusted authorities are significantly reduced; with the Governance actors not having to worry about the functional requirements of ***Data Sovereignty*** and ***Secure Data Exchange*** as they are fully enforced by the MPC technology**.**

The updated list of categories and codes representing the effect of MPC technology on the architecture of the data marketplace platform is list below (additions in ***green*** and deletions in ***red***):

- ***T7:*** *MPC Incorporation into the Data Marketplace Platform*

- ***C1:*** *Enables Decentralized Design*
- ***C2:*** *Powers Data Exchange Service*
  - ***C21:*** *Safe-DEED Component*
    - ***C211:*** *Safe-DEED Primitives*
    - ***C212:*** *Safe-DEED Network*
  - ***C22:*** *Moves Data Analysis to Data Exchange Service*
    - ***C222:*** *Data Analysis service changes to Data Analytics AppStore*
- ***C3:*** *Reduces the burden of Governance*
  - ***C31:*** *Enables Data Sovereignty technologically*
  - ***C22:*** *Enables Secure Data Exchange technologically*
  - ***C23:*** *Moves Data Governance to Data Provider's site*
  - ***C24:*** *Changes Data Inventory to Metadata Inventory*
- ***C3:*** *Enables Security-by-Design*
  - ***C31:*** *No need for Key Management*

These changes result into the updated *high-level architecture reflecting MPC incorporation*, the ***Post-MPC Data Marketplace Platform 2.0*** as illustrated in **Figure 19** (changes highlighted in ***yellow***).

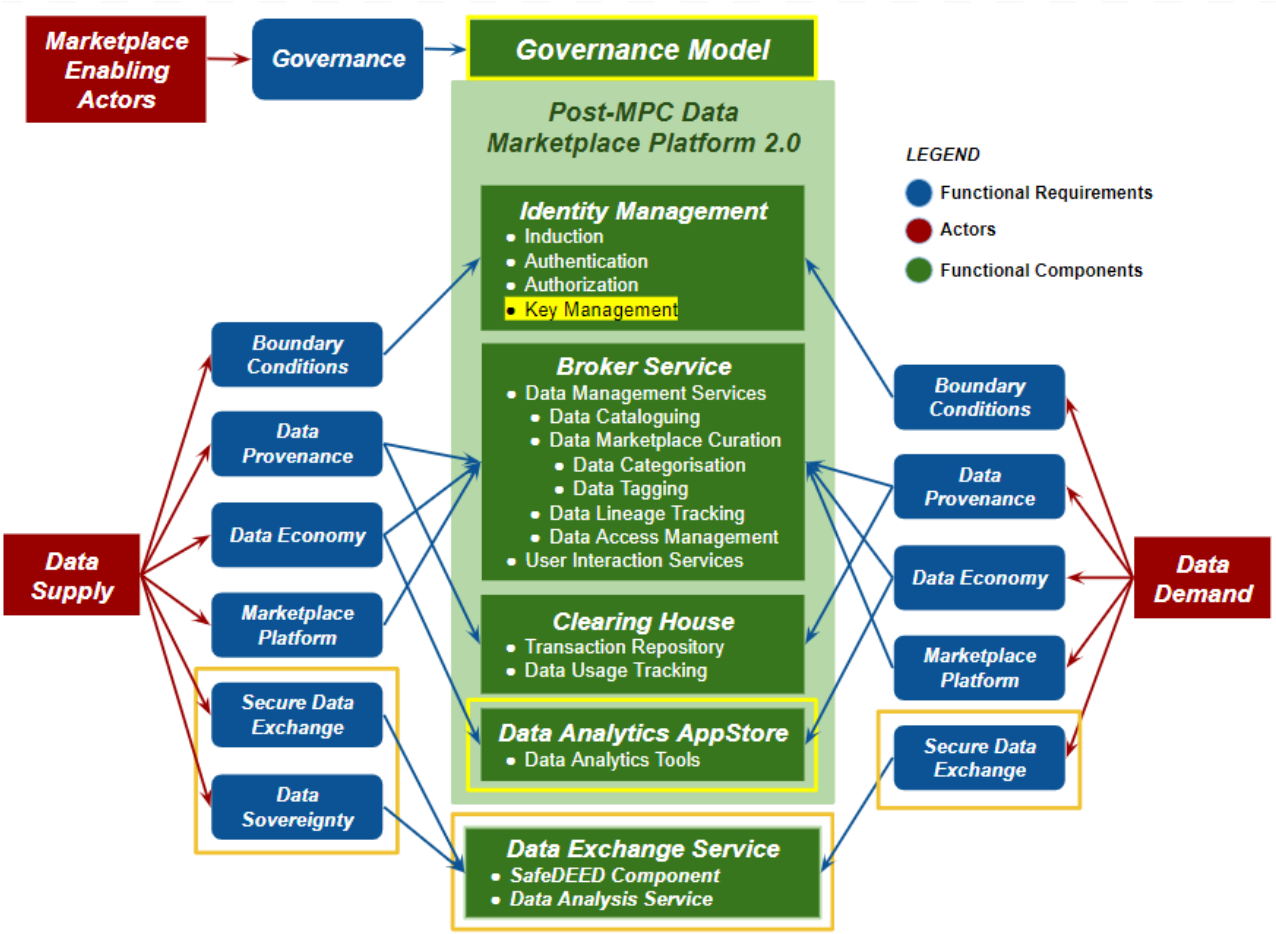

Figure 19: Post-MPC Data Marketplace Platform 2.0

6.4 Validation of Pre-MPC Threat Model 1.0

The topics and their theoretical concepts associated with the research focus, *RF3* are validated here. The conceptual model under consideration here is the *Pre-MPC Threat Model 1.0* built in Chapter 3. This topic was intended to be validated mainly from the cybersecurity expert, *E3* which we did. However, it turned out that expert *E1* also had expertise in this area and *E1* was kind enough to give his insights here. The advantage of having *E1* onboard for this topic was that *E1* is an expert in data marketplaces

and hence, we got valid insights related to how to approach the threat aspects of data marketplaces in addition to the process of threat modelling in general. The following 2 topics are validated in the upcoming subsections.

- ***T8:*** *HLTM Framework*
- ***T9***: *Threat Landscape of the Data Marketplaces*

6.4.1 HLTM Framework

HLTM framework is a new framework developed by us for the context of high-level threat modelling, and since, threat modelling is a crucial aspect of our research objective, the topic, *T8: HLTM Framework* was included as part of validation activity. The initial list of categories and codes in this topic derived from Chapter 3 were:

- ***T8:*** *HLTM Framework*
- ***C1:*** *Context of Threat Modelling*
  - ***C11:*** *Scope*
    - ***C111:*** *At the level of business functions*
  - ***C12:*** *Approach*
    - ***C121:*** *Asset-Centric*
  - ***C13:*** *Purpose*
    - ***C131:*** *Risk Framing – Risk Management*
  - ***C14:*** *Context Statement*
    - ***C141:*** *"to establish the assets associated with the business functions of each functional component of the high-level architecture of a technological entity and later, assume a system specification on which applicable cyberattack vectors (described at a high-level) can be identified"*
- ***C2:*** *Type of Threat Model*
  - ***C21:*** *High-Level Threat Model*
- ***C3****: Constructs*
  - ***C31:*** *Functional Component*
  - ***C32:*** *Business Function*
  - ***C33:*** *IT System Asset*
    - ***C331:*** *Primary Asset*
    - ***C332:*** *Supporting Asset*
  - ***C34:*** *Threat*
    - ***C341:*** *Cyber Attack Vector*
    - ***C342:*** *System Failures*
  - ***C35: CIA Violated?***
    - ***C351:*** *Confidentiality*
    - ***C352:*** *Integrity*
    - ***C353:*** *Availability*
  - ***C36:*** *Business Consequence*
  - ***C37:*** *Mitigation Technique*
- ***C4:*** *Threat Landscape*
  - ***C41:*** *Threat*
  - ***C42:*** *Business Consequence*
- ***C5:*** *Limitation*
  - ***C51:*** *Baseline Overview*

6.4.1.1 Results & Analysis

When asked about our process of threat modelling using the HTLM framework and the threat model, *E1* reflects that the threat modelling here "*assumes certain implicit architecture*". So, "***the threat model could change if you take a different architectural design***". The *implicit architectural decisions* taken in the *component and business function* construct of how the assets are handled in a data marketplace are an assumption. If the component and business function are implemented architecturally in a different

way other than our assumption, then the threat model does not apply. *E1* basically suggests that the threat model will be valid only if there is a defined and detailed underlying architecture. Furthermore, *E1* says that the threat model is valid only to that specific architecture. However, *E1* says that our method is ***fine to obtain baseline threats to the focal entity and hence, its baseline security requirements***. But again, *E1* criticizes our threat model to be a "***low-level threat model***" containing threats to a lower level architecture of the components of the data marketplace which will be addressed by the chosen mitigation techniques. But the threats crucial to the data marketplaces are the ones at the ***higher-level*** like ***"data leakage"*** which are "***difficult to identify***" and "***more complicated***" for our chosen mitigation techniques to prevent; and hence, need "***special mechanisms to mitigate***". *E1* further gave few examples of these higher-level threats to the data marketplaces which will be discussed when dealing with the associated threats. *E1* suggests that in order to find ***higher-level threats,*** we should understand the main business logic of the data marketplace which is handling data, and hence, we should focus on threats associated with "***data sensitivity***".

E3 reflected overall that the framework and the threat model were relevant and strong compared to the industry standards. However, E3 suggested a few relevant aspects. *E3 suggests that "when looking at the business functions" to do security assessment, we are supposed to consider the* ***processes or procedures****, the* ***requirements*** *towards cybersecurity and how these requirements are enforced within an organization. E3* recommended including ***vulnerabilities*** as a construct as it is the only missing cybersecurity in the framework. About the threat model, *E3* remarked that the threat model is good and comprehensive and suggested few more threats like *system failure, server unavailability, malicious insider* et cetera which again belong to the category of "***low-level threats***" of *E1*. *E3* further suggests including threats like ***regulatory, environmental, mismanagement of personally identifiable information*** et cetera to the threat model saying that these are just as relevant as ***IT threats***. Apart from that, *E3* was fine with the framework reflecting that the framework would give ***a generic direction*** towards the security of the focal entity. But *E3* suggested that after generating a high-level threat model, it is necessary to do ***second round of security assessment customised to the specification of the focal entity***. Consequently, *E3* suggested having actual "***architectural concepts of data marketplaces***" in place to "***find valid threats***" echoing the same insight as that of ***E1***. When asked about the significance of high-level threat models generally, ***E3*** echoed our view by saying "*it is a good start to have a set of high-level threats applicable to a type of focal entity*".

6.4.1.2 Conclusions

The general insight about the HLTM framework and its resulting threat model is that it only represents the starting direction towards the security design of the focal entity. Both *E1* and *E3* reflect this through their "*implicit architecture*" comment (as a result of which it cannot be generalized but only represents a baseline overview) and "*good start to have a set of high-level threats*"; echoing the limitation of our framework that it only provides a baseline security overview which we already have established in the Chapter 3.

Second crucial insight was to go beyond the *IT threats* (*cyber threats*) which is echoed by both the experts. However, our context clearly mentioned the reasoning for this choice that the cyberattack vectors represent the tactic-level description of the technological platform. However, the insight of "***higher-level threats"*** by *E1* is interesting. He basically means that the threats being focused here are cyber threats operating at the system-level business functions. These threats can be overcome easily through mitigation techniques. However, the threats which exist at a relatively higher-level than the systems' business functions are crucial for the data marketplaces as these threats can disbar the business logic of the data marketplaces are very complex to solve. The example of "data sensitivity" reflects the same that even though the whole system is 100% secure, if an authorized customer behaves malicious where he misuses the data (by leaking it or using for means other than the ones in the contract) which was legally purchased from the data owner. In that case, the mitigation technique could not do anything as everything is working fine but the problem lies in the fundamental business logic of the entity. In the case of data marketplaces, *E1* suggested that the fundamental business logic is the handling of the data and the sensitive nature of protecting it. This relates to the *challenges of commoditizing* data that we discussed in Chapter 2. Hence, the nature of data needs to be studied and that knowledge should be applied in the contexts of data marketplaces and gauge how things can go wrong and how that can

impact the data marketplaces as organizations. Basically, the data marketplaces should be looked at as business entities than just technological platforms to find the threats that are crucial for the functioning of the data marketplaces i.e. the threats which actually reflect the threat landscape of the data marketplaces. *E1* further remarks that these threats are "***difficult to identify***" and are "***more complicated***" for our chosen mitigation techniques to prevent; and hence, need "***special mechanisms to mitigate***". We term these threats as ***"business threats***" as they affect the business logic which is a far higher level than the high-level cyber threats to the business functions of the individual information system components within the entity (which is what performed in HLA framework). This insight was incorporated into the from the NCGI Apex Classification of threat models in the form of a new category of "***business threat models***" which was added at the level beyond the other threat models which only deal only with cyber threats, see **Figure 20**. Furthermore, in the light of this insight, it was decided to start referring to our focal threats as cyber threats explicitly as they are different from the business threats as mentioned here.

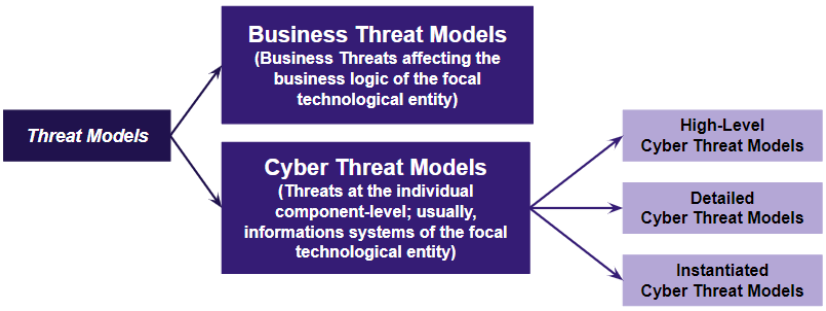

Figure 20: Threat Model Taxonomy 2.0

However, the business consequence construct reflects the effect of the cyber threats to the business functions or the whole focal entity. The latter aspect signifies that the consequence is established not just at the level of information systems or the low-level business functions but at the higher-level of the whole organizations. Hence, this construct reflects the concept of the *business threats* as introduced previously. As a result, the construct can be directly renamed to be called "***business threats***". Furthermore, the insight that these business threats actually represent the threat landscape of the focal entity coincided with our conceptualization of threat landscape as we had included business consequence as well. It made sense to incorporate threat construct given we were aiming to get a baseline threat landscape. Now that it is established that those threats can be easily overcome by mitigation technologies, it no longer reflects the actual threat landscape of the focal entity. Hence, we shall now refine the conceptualization to include only the construct of *business threats* alone. However, the cyber threats still contribute here indirectly as they can influence business threats into manifesting. The threat of *"mismanagement of personal identifiable information"* suggested by *E3* reflects this scenario where the breach of PII can affect the business logic of the focal entity is data security was its business logic. However, this scenario applies in the presence of a detailed technical architecture as that guarantees the reflection of actual threat landscape. The refined conceptualization is illustrated in **Figure 21**.

Figure 21: Conceptualization of the Threat Landscape 2.0

Related to the framework, we rename it to ***High-Level Cyber Threat Modelling (HLCTM) framework*** to reflect the cyber aspect discussed earlier. Furthermore, we add the construct of ***vulnerability*** into the framework. Vulnerability is a design flaw that exist in the system under focus which can be exploited by cyberattacks. In the context of the availability of a specific architecture, vulnerability is a relevant construct and hence qualifies to be added into the framework. The resulting cyber threat model will be specifically valid to the architecture under consideration. For high-level cyber threat modelling, however, the construct can be ignored. This move increases the flexibility of the already flexible cyber threat modelling framework, see **Figure 22**.

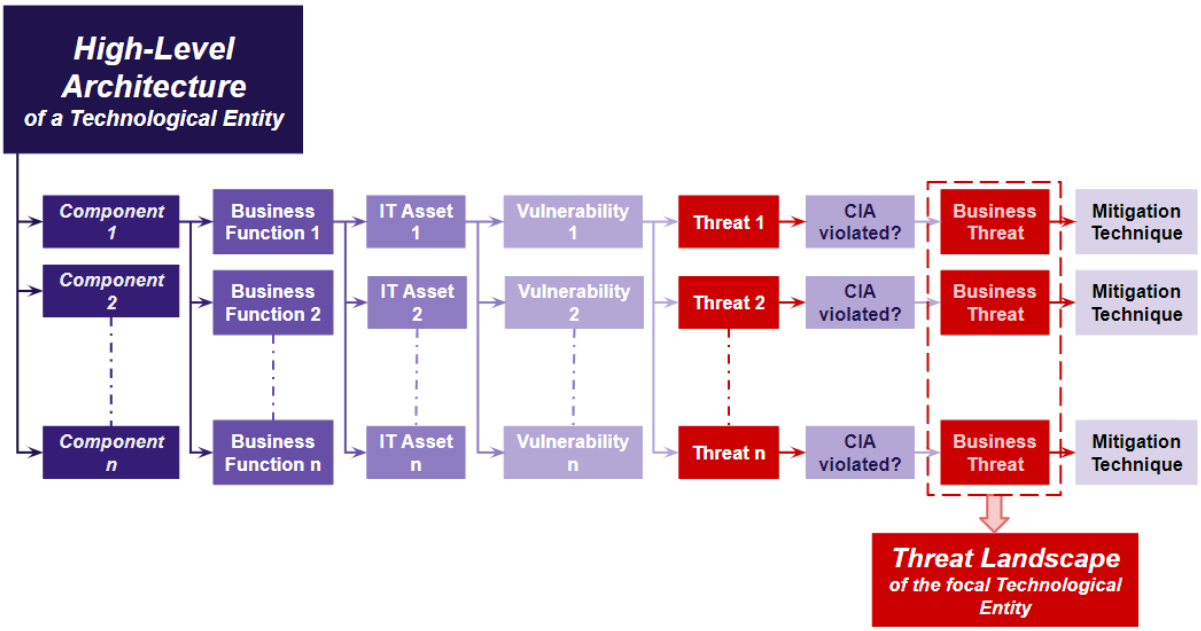

Figure 22: High-Level Cyber Threat Modelling (HLCTM) Framework

The framework was not used again to find the business threats as in the framework, only the cyber threats drive the business threats and are applicable to the detailed technical architectures. However, in our case, this does not apply as our architecture is still high-level. To identify the actual high-level business threats specific to data marketplaces, it is advised either to explore the concept of ***data sensitivity*** and understand the threats around it; or to understand the business logic of a data marketplace by carrying out a case study of a real-life data marketplace and then, identify the business threats to that data marketplace which can then be generalized to all the data marketplaces. However, we used the interviews with the experts to do so which will be discussed in the next subsection.
The updated list of categories and codes associated with *T8: HLCTM Framework* is listed below with the modifications highlighted (additions in *green* and deletions in *red*):

- ***T8:** HLCTM Framework*
- ***C1:** Context of Cyber Threat Modelling*
  - ***C11:** Scope*
    - ***C111:** At the level of business functions of information systems*
  - ***C12:** Approach*
    - ***C121:** Asset-Centric*
  - ***C13:** Purpose*
    - ***C131:** Risk Framing – Risk Management*
  - ***C14:** Context Statement*
    - ***C141:** "to establish the assets associated with the business functions of each functional component of the high-level architecture of a technological entity and later, assume a system specification on which applicable cyberattack vectors (described at a high-level) can be identified"*
- ***C2:** Type of Cyber Threat Model*
  - ***C21:** High-Level Cyber Threat Model*
- ***C3**: Constructs*
  - ***C31:** Functional Component*
  - ***C32:** Business Function*
  - ***C33:** IT System Asset*
    - ***C331:** Primary Asset*
    - ***C332:** Supporting Asset*
  - ***C34:** Vulnerability*
  - ***C35:** Threat*

        - ***C351:*** *Cyber Attack Vector*
        - ***C352:*** *System Failures*
    - ***C36: CIA Violated?***
        - ***C361:*** *Confidentiality*
        - ***C362:*** *Integrity*
        - ***C363:*** *Availability*
    - ***C37:*** *Business Threat*
    - ***C38:*** *Mitigation Technique*
- ***C4:*** *Threat Landscape*
    - ***C41:*** *Business Threat*
- ***C5:*** *Limitation*
    - ***C51:*** *Baseline Overview*

6.4.2 Threat Landscape of the Data Marketplaces

Following the discussion from the previous subsection, it was deduced that the *High-Level Threat Model* from the Chapter 4 does not reflect the actual threat landscape of the data marketplace platform but only a baseline overview which does not actually contribute towards understanding the effect of MPC technology on the data marketplace platforms. Hence, although a good and comprehensive cyber threat model, it was discarded to be invalid for our objective. The same was seconded by the expert, *E1* that the threat model does not represent the threat landscape specific to data marketplace platforms. Even though, the threat model had business consequence construct which relates to the actual threat landscape, it was discarded as the value of the business consequence were driven from the baseline cyber threats. Hence, we start from the scratch here with no *Pre-MPC Threat Model.*
However, we got rich and appropriate insights from the experts about the so-called *business threats* which prevail specifically for the business logic of the data marketplace platforms. We gained these insights from the experts *E1* and *E2* who are well versed in the field of data marketplaces and hence, reflected well on the subject of the threats associated with them. In addition to this, *E4* also contributed with a threat scenario which applies here. We present the same insights and analyze them to generate the valid list of threats which do reflect the actual threat landscape of the data marketplace platforms, reflecting the *Pre-MPC Threat Model 2.0*. Furthermore, the new list of threats is generated by conducting ***Open Coding*** through which codes are generated form the data without any initial list of categories and codes. As a result, we end up with a fresh list of categories and codes straight from the data.

6.4.2.1 Results & Analysis

With respect to the threats associated with the data marketplace platforms, the experts talked mostly about high-level business threats while giving threat scenario examples for few threats. In this subsection, we shall present the threats in a qualitative way as described by the experts and codify accordingly.
E1 tells that the issues that are currently crucial to the data marketplaces are not the attacks from external entities, but the internal problems within the data marketplace. According to *E1*, these are the issues being worked and researched on by the industry rather than the cyberattacks. *E1* gives examples of these issues and they are quoted here with their respective labels that we have coded.

- *"once the consumer gets the data, nothing stops the consumer to do whatever he wants with the data".* We reduce this statement into 2 codes, ***Loss of Control over Data*** and ***Data Leakage;*** and assign it to the category of ***Threats.*** Furthermore, relating about mitigating, *E1* says*, "this comes back to governance".* We code this with label, ***Governance*** and assign it to the category of ***Mitigation Techniques.***
- *"the internal actors have to work solely on the trust over the parties".* We code this with label, **Trust Issues** and assign it to the category of ***Threats.***

- In decentralized design, where the access of data is given over an encrypted channel, on the receiving end, *E1* states that the algorithm at the receiving end can copy the data by saying, "*the container with an algorithm which need to decrypt the data. However, the container is going to copy the data somewhere. So, the basic security is broken. You have to do much to work to enforce which involves governance*". Evidently, the ***threat*** aspect of this statement can be coded into ***Loss of Control over Data*** and ***Data Leakage***; and the ***mitigation technique*** to be ***Governance***.
- "*the data providers are going to trust you with his credentials, and some would not trust and would not give the keys"**. Threats > Trust Issues.***
- "*You have higher level threat models like of data leakage. For example, you're sharing data under some contract which restricts its usage and so on. The threat models where the data can be correlated with another data set which can lead to some leakage. For example, you try to anonymize the data, for example, by removing some items like personally identifiable information. One of the threat models is that this anonymized data can be correlated back to the identities if it is combined with other appropriate datasets. So, these are the high level threat model cases that still need to be addressed in the setting of a data marketplace."* We code this account with the label, ***Data Leakage by Back Correlation*** and categorize it to the ***Threats***.
- *"are high-level like data leaks. For example, if you provide a database of all the people in the Netherlands, and you try to anonymize it, and you say you can use this data, then there exists a threat model that someone would get this anonymized database and correlate it to the identities. The threat model referred runs as an application which correlates the anonymized data to the identities."* Same case as the previous account. Hence, ***Threats > Data Leakage by Back Correlation.***
- When talking about how to identify these high-level business threats, "*I came across them mostly from talking about data sensitivity"***.** Hence, we code this statement as ***Data Sensitivity*** and categorize as broad one in ***Threats.***
- *"And with a lot of AI being done now, there's all this back correlation of census data".* This is another case of the same code, ***Data Leakage by Back Correlation***.
- *"The problem with sensitive data is that it can relate logically. So architecturally, you have everything secure. But the algorithm that's being applied on the data can itself be a threat as it can cause data leak. i.e. if the data is not properly anonymized, then the leaked data itself can be sensitive even without back correlation. So, in this example, even how to anonymize can be a big issue."* This is a simple case of data leakage and hence the code label is ***Data Leakage***. The statement can also be coded to ***Data Sensitivity***.
- "*for example, the MRI images. You can say the MRI images itself can be processed and with tracking preferences, determine that the MRI image itself is already identifying people because it becomes like a fingerprint. Although it's anonymized and it doesn't mean anything to get the MRI images, but by correlating it with image processing and other dataset, the sensitive data can find the home it belongs."* This is again an example for a case of the same codes, ***Data Leakage by Back Correlation*** and ***Data Sensitivity.***
- "*I don't think there's a way to mitigate these risks to hundred percent. That's where you come back to governance issue because eventually, when there is some data leak, you have the auditors and everything that you can litigate legally."* This statement clearly belongs to the code, ***Governance*** of the category ***Mitigation Techniques.***

Now, moving on to our next expert, *E2* provided some additional inputs towards the subject of threats associated with the data marketplaces. *E2*'s insights are quoted below and are coded and categorized accordingly.

- *"As long as somebody has access to the data, they can write a function on it, that copies the data, and then, subsequently misuse it. That is their intention".* This statement can be related to the code of ***Loss of Control over Data*** and ***Data Leakage;*** and categorize it to ***Threats***.
- "they will lose control over their data and that the data will be out in the wild, even if it is behind the paywall. Somebody else will pay for it and then they will release it. They give the example of obviously, movies or whatever. They are all behind the paywall, and then somehow, they all

ended up on some BitTorrent site". ***Threats > Loss of Control over Data*** and ***Threats > Data Leakage***

- "*in production and manufacturing, producing data from the machines has the potential danger of a competitor reverse engineering their processes. For instance, it can be like, they have a special process that they produce some plastic at a certain temperature, which makes it better or more stable. And then if they release sensor data from the machines about energy consumption and operation times, then based on the energy consumption, perhaps the competitor will be able to determine the temperature they're using in the process. This is an example in the industry and manufacturing".* Classically, this instance appends to the code of ***Data Leakage.*** However, this account brings about a new code with label, ***Loss of Competitive Advantage for Data Actors*** caused by the disclosure of proprietary information. This is categorized to ***Threats.***
- *"In all other sectors like Banking, Telecom or the Health, of course, the problem is with the regulations. They are afraid at some point that the data will be deanonymized and therefore, they will be facing fines for having released personally identifiable information."* This statement can be coded with label, ***Regulatory Threats*** and ***Data Leakage by Back Correlation*** (deanonymization is done through back correlation of data). Additionally, it is categorized into ***Threats.***

In addition to these, *E4* suggested a threat scenario which applies in this situation. This is shown in a qualitative way below and is coded accordingly.

- *"if you are a malicious actor, and you use the marketplace, and you do computations with everyone; but you always just make up all the data, then it doesn't look too good from the perspective of the marketplace".* This statement is reduced to a new code, ***Induction of Malicious Data Actor*** and is categorized under ***Threats***

Ultimately, under the topic of Threat landscape of the Data Marketplaces, we shall have 2 categories: ***Threats*** and ***Mitigations***. In the category of *Threats,* we end up the codes *Loss of Control over Data*, *Trust Issues*, *Data Leakage*, *Data Leakage due to back correlation of data*, *Loss of Competitive Advantage*, *Regulatory Threats*, *Data Sensitivity* and *Induction of Malicious Data Actor*. On the other side, the category of *Mitigation Techniques* consists of only one code, *Governance.* The final codes and categories are listed in **Table 21**.

Table 21: Updated Categories and Codes and their number of references by Experts

Category	*Code*	*Suggested By*	*Instances mentioned*	*Total No. of instances*
Threats	*Loss of Control over Data*	*E1*	2	**4**
		E2	2	
	Trust Issues	*E1*	2	**2**
	Data Leakage	*E1*	3	**6**
		E2	3	
	Data Leakage by Back Correlation	*E1*	4	**5**
		E2	1	
	Loss of Competitive Advantage for Data Actors	*E2*	1	**1**
	Regulatory Threats	*E2*	1	**1**
	Data Sensitivity	*E1*	3	**3**
	Induction of Malicious Data Actor	*E4*	1	**1**
Mitigation Techniques	*Governance*	*E1*	2	**2**

6.4.2.2 Conclusions

Here, we shall convert the codes and categories obtained from the previous subsection into the Business Threat Model in lines with the objective of this research. The Business Threat Model is illustrated in the **Table 22**. The constructs used here in this threat model are: Business Threat, Threat Description, Threat Experiencing Actor and Mitigation Technique. The threats are described as appropriate to the *Pre-MPC Data Marketplace Platform 2.0*. However, different interpretations of these threats apply in all the designs of the data marketplace platforms. The Business Threat Model, also representing ***Pre-MPC Threat Model 2.0***. represents the ***actual threat landscape*** of the data marketplace platform.

Table 22: Pre-MPC Threat Model 2.0

Business Threat	*Threat Description*	*Affected Actor*	*Mitigation Technique*
Loss of Control over Data	The threat comprises of instances where once the data is transacted and is away from the data owner, the data can be exploited to do anything. It can be used for malicious activities, or it can be resold to some other actor or it can simply be copied and released over internet. Since the data owner legally owns the data and licenses it to the consumer, then if that data is used by the consumer for malicious activities, then even the data owner will be held legally liable for that malicious act since the data he legally owns was used there.	Data Owner	***Governance Model***
Trust Issues	Since the actors are expected to participate based only on the trust towards the marketplace authority and the other data actors and since, there is no tangible way of proving the trust mechanism in place and also unavailability of any technological way of enforcing trust, the data actors may not participate in the data marketplaces as they don't trust somebody else with their data. This turns out to be a threat to the Marketplace provider.	Marketplace Provider	
Data Leakage	This threat is a straight forward one where the data being transacted gets used by the involved data actors in a way that was not intended by the data owner in the terms of the contract. So, the data is being used as not intended which is a threat to the data owner and also to the supply side as they are also involved in processing the data to be transacted.	Supply Side Actors	
Data Leakage by Back Correlation	A special kind of Data Leakage threat where the data with personally identifiable information (PII) is anonymized and transacted to the consumer; and then, the PII can be extracted from the data either because of faulty anonymization or by combining it with other auxiliary data sets, and eventually correlating it back to the original PII.	Supply Side Actors	
Loss of Competitive Advantage for Data Actors	This is a different kind of threat resulting out of data leakage threat where, from the shared data, the receiving actor learns some proprietary information about the data owner selling the data or the supply side actors involved in the business process of the data transaction. The case of back correlation or combining with other data by the receiving actor can result in the loss of competitive advantage to any of the applicable supply side actors.	Supply Side Actors	
Regulatory Threats	This is the legal aspect of all the threats covered here. The threats discussed till now can result in regulatory threats for various reasons like violation of the terms in the contract, violation of privacy et cetera. The logistics of how exactly this threat apply is dependent on specific cases.	All Actors	
Data Sensitivity	This is a broader threat which relates to the unique characteristics of data that makes it challenging to commodities it as discussed earlier. It can be stated that all the threats associated with data are due to this broad threat of Data Sensitivity. All the threats dealt before this here can be stated as specific cases resulted because of the sensitive nature of the data.	All Actors	
Induction of Malicious Data Actor	This is a generic threat to any marketplace where a malicious actor is inducted into the data marketplace as a legitimate costumer. The data actor can be on the platform to exploit the services which is a risk to the data marketplace. This data actor can provide bad data for the computations, thereby generating invalid results for the fellow actors.	All Actors	

The final list of the codes in the topic *T9* are as follows:

- ***T9: Threat Landscape of the Data Marketplaces***
- ***C1:*** *Loss of Control over Data*
- ***C2:*** *Trust Issues*
- ***C3:*** *Data Leakage*
- ***C4:*** *Data Leakage by Back Correlation*
- ***C5:*** *Loss of Competitive Advantage for Data Actors*
- ***C6:*** *Regulatory Threats*
- ***C7:*** *Data Sensitivity*
- ***C8:*** *Induction of Malicious Data Actor*

Coming to the Mitigation Technique aspect of the Business Threat Model, the experts feel that the technology is not mature enough to address the issue of ***Data Sensitivity*** and its ramifications (other threats resulting out of ***data sensitivity***); and hence, cannot enforce the functionalities of the data marketplace platform in a comprehensive way. They still feel that it is a collective coordination between the technology, regulation and the actors involved complementing into a ***Governance model*** which enforces every functional requirement along with the security. However, it goes without saying that a 100% security is never possible, and the threats are never mitigated or eliminated but only minimized. The effective enforcement of the ***Governance Model*** is the solution towards minimizing the threat influence on the data marketplaces.

6.5 Validation of the Post-MPC Threat Model 1.0

The topics and corresponding theoretical concepts of the research focus, *RF4* are validated here. The artefact under consideration here is the *Post-Data Marketplace Platform 1.0* from Chapter 5. The intention here is to investigate the effect of MPC on the threats associated with the data marketplaces. Consequently, the topic validated here is,

- ***T10***: *Effect of MPC Incorporation on the Threat Landscape of the Data Marketplace Platform*

6.5.1 Effect of MPC Incorporation on the Threat Landscape

This topic represents the second of the two flagship conceptual models contributing towards our research objective as validation of this topic contributes towards the understanding of the implication of MPC technology to the threat landscape of the data marketplaces. As the high-level threat model from Chapter 4 was deduced to be invalid to our objective as it does not reflect the threat landscape of the data marketplaces, the same thing applies to the *Post-MPC Threat Model 1.0*. As a result, there is not validation in this section. But the final conceptual model of this research ***Post-MPC Threat Model 2.0*** is developed with the insights of experts, *E1* and *E4*.

6.5.1.1 Results & Analysis

E1 suggests that even after MPC is in place, the threat of trust issues exist in the sense that since the governance authority is eliminated and the trust surrounding data is totally handled by MPC technology, the data owner might find it difficult to trust the other data actors in the absence of the governance authority. Hence, *E1* recommends some form of governance to tackle these threats. This issue relates to the business threat of **trust issues** in the Pre-MPC threat model 2.0.
On this subject, *E4* remarks that with MPC in place and the absence of governance authority, the threat of malicious data actor increases as the trust mechanism is maintained by technology and the malicious data actor will get away with providing faulty data and using the service and resources of the data marketplace. However, the threat of malicious data actor impacts severely on the other data actors while only causing reputation loss to the marketplace provider. But *E4* reflects that this may escalate if there are more malicious data actors than legitimate and honest data actors. In that case, *E4* reflects if contracts between the data actor and the marketplace provider are set up and hold accountable legally if the data actors behave maliciously. This relates to business threat of ***malicious data actor*** in the ***Pre-MPC threat model 2.0***.

6.5.1.2 Conclusions

The Post-Threat Model 2.0 comprising of the business threats which prevail for the data marketplaces even after the incorporation of MPC technology is generated here. Firstly, the effect of MPC technology on each business threat in the ***Pre-MPC Threat Model 2.0*** is discussed and the ones which will be mitigated by MPC are filtered. At the same time, the business threats post-MPC as identified from the insights of the experts to finally obtain the ***Post-MPC Threat Model 2.0.***

- ***Loss of Control over Data:*** Since there is no actual transfer of the data from the data owner to the other actors and the fact that the data resides at the site of the data owner and provisioned remotely through MPC protocol, the business threat of loss of control over data does not apply anymore. However, the terms of how the data is provisioned for the MPC protocol by the data owner should be stipulated over the contract and there should be governance model to enforce this.
- ***Trust Issues:*** This business threat transforms into a different case of trust issue where the data actors find it difficult to trust other data actors in a technological setting with the ***absence of the authority***. To tackle this, since MPC technology ensures the enforcement of trust surrounding the data, there should be a ***governance model*** to handle the trust associated with the rest of the aspects of the data marketplaces.

- ***Data Leakage:*** On the assumption that MPC protocol works efficiently and effectively, since there is no actual transfer of the data between the actors, the business threat of data leakage does not apply any more.
- ***Data Leakage by Back Correlation:*** Here, the same thing applies as the previous threat and hence, even this business threat is no more applicable.
- ***Loss of Competitive Advantage for Data Actors:*** The same reason as the previous two business threats apply here too. However, this can depend on the function or the data analysis service in the MPC protocol as the receiving actor can further analyze the computation result by combining it with other auxiliary data or reverse engineering et cetera. With the effective and efficient execution of the MPC protocol, this business threat does not apply.
- ***Regulatory Threats:*** The same reasoning as the previous business threat applies here where in the intended functioning of MPC protocol, the business threat does not apply. However, the business threat can apply in extreme cases of data leakage due to faulty execution of MPC protocol.
- ***Data Sensitivity:*** This is the business threat that MPC Protocol is specifically tackling, addressing and mitigating. Since MPC protocol ensure the functional requirements associated with data sensitivity, ***Data Sovereignty*** and ***Secure Data Exchange***, the business threat of data sensitivity does not apply anymore.
- ***Induction of Malicious Data Actor:*** With MPC protocol in place and the absence of governance model, the business threat of malicious data actor increases as the trust mechanism is maintained only by the technology. The business threat can escalate to a detrimental level when the number of malicious data actors present on the platform exceed the number of legitimate and honest data actors. Hence, this business threat prevails as it affects the functional components of ***Boundary Conditions*** and ***Secure Data Exchange*** and can be addressed through a ***stricter induction*** of data actors as part of ***governance model*** to enforce non-data sensitivity related trust governance while letting MPC technology to enforce data sensitivity related trust maintenance. Furthermore, a more sophisticated contract management enforcing the terms of the contracts between the data actors and marketplace provider can be incorporated. Perhaps, upgrade the contract management with Blockchain Technology.

This brings us to the end of our discussion about the effect of MPC technology on the threat landscape of the data marketplaces. During this discussion, we came across the issue of the effective and efficient execution of the MPC protocol where if this is compromised, all the business threats mitigated by the MPC technology shall return and apply again. Hence, we shall include this as a business threat of ***Faulty Execution of MPC protocol*** in the threat model. With this business threat, the uncertainty involved with the business threats of loss of competitive advantage and regulatory threats is addressed. *Faulty execution of MPC Protocol* can be mitigated by employing auditing authority who can carry out auditing of the MPC protocol and its associated processes, essentially qualifying to the mitigation technique of ***Governance Model*** which includes ***MPC process auditing***. The resulting threat model is the ***Post-MPC Threat Model 2.0*** and is shown in the **Table 23**.

Table 23: Post-MPC Threat Model 2.0

Business Threat	*Threat Description*	*Affected Actor*	*Mitigation Technique*
Trust Issues	The threat of data actors not participating in the data marketplace as the data actors find it difficult to trust a technological setting of just MPC protocol to handle their valuable commercial data	Marketplace provider	***Governance Model with MPC Process Auditing***
Induction of Malicious Data Actor	The threat of inducting a malicious actor into the data marketplace as a legitimate costumer. With MPC protocol in place and the absence of governance model, the threat of malicious data actor increases as the trust mechanism is maintained only by the technology. The threat can escalate to a detrimental level when the number of malicious data actors present on the platform exceed the number of legitimate and honest data actors; affecting the functional components of ***Boundary Conditions*** and ***Secure Data Exchange***	All actors	
Faulty Execution of MPC Protocol	The compromise of the intended (effective and efficient) execution of MPC protocol which can result in the return of all the threats associated with data sensitivity.	All actors	

In the ***Post-MPC Threat Model 2.0,*** it can be seen that MPC technology eliminates the business threats of ***Loss of Control over Data, Data Leakage, Data Leakage by Back Correlation, Loss of Competitive Advantage for Data Actors, Regulatory Threats*** and ***Data Sensitivity.*** So, basically, MPC eliminates the business threats associated with the issue of ***Data Sensitivity*** and its ramifications; and MPC does so still in a ***Security-by-Design*** way as mentioned in Chapter 5; thus, reducing the burden of the ***Governance Model*** on its technological front.

This page has been intentionally left blank.

7 Discussion

In this chapter, we discuss our findings and relate them to the initial research objectives. Furthermore, we identify the correct incentive structures and models, i.e., advantages of the adoption of MPC technology for distributed data marketplaces, to facilitate their adoption in the European economic area.

7.1 Architectural Implication of MPC to the Data Marketplaces

The conceptual model representing the first half of the theoretical framework is described here. The conceptual model reflects the implication of the MPC technology to the architectural aspects of the data marketplaces by explicating the difference between *Pre-MPC Data Marketplace Platform 2.0* and *Post-MPC Data Marketplace Platform 2.0*; further generalizing the same for the business species of data marketplaces.

Enables Data Trading in a Confidentiality-Preserving and Privacy-Preserving way
MPC technology could enable data trading and data sharing to happen in a confidentiality-preserving and privacy-preserving way where the data owners do not have to transfer the physical data to the receiver. Instead, the data sharing process is converted into a *cryptographic protocol* through which only the result of the computation on the data (or the union of data in case of multiple parties) is shared with the dedicated receiver(s) with the actual input data not revealed to any of the parties involved in the transaction. Since the transfer of physical data is not present, MPC improves the business potential of data trading for all the actors involved in the data marketplace ecosystem. Additionally, the need of anonymization for data owners becomes irrelevant because of privacy-preserving nature of the MPC protocols.

Transforms the Data Exchange Service to ensure Secure Data Exchange
The traditional *Data Exchange Service* which was assumed to an SSH encryption-based communication channel, associated with a vulnerable and costly *key management system* which involves physical data being encrypted and sent over the channel to the receiver who can obtain the decrypted form of the physical data. If the receiver uses the received data for supposes other than the terms of the transaction, there is no way of knowing that because of the *cooperative* nature of data (that it can be replicated at negligible cost and used simultaneously). With MPC technology, it could be transformed into a safer and sophisticated MPC protocol which not only eliminates the *key management system* for the data marketplaces but also enforces the functional requirement of *Secure Data Exchange* for the data marketplace ecosystem.

Enables Data Sovereignty for the Data Marketplace to be truly-decentralized
Because of MPC technology, the physical data could no longer be transferred to other entities. Instead, the data owners could hold the data with themselves and could provision it to the dedicated receivers like data aggregators or data consumers et cetera with the MPC protocol which eliminates the need for a governance authority to overlook this process. Through this, the MPC protocol enforces the functional requirement of *Data Sovereignty* for the data owners thereby overcoming the challenges of commoditizing data: *Protection Regime* and *Quality Control*. This property enables the data marketplace to be *truly decentralized* which makes it a trustworthy platform for data trading.

Supports Data Analysis Services
MPC protocols enable data analysis service as part of their protocols; thus, could lighten the data marketplace platform of the responsibility and infrastructure of data analysis services. It also enables the data analysis service providers to operate trust-free, independent of the data marketplace platform.

Reduces the burden on Governance Model
Since MPC technology alone enforces the functional requirements of *Data Sovereignty* and *Secure Data Exchange* technologically, its incorporation eases the responsibilities of the *Marketplace Enabling actors* with respect to *Data Governance.* However, they still must look after *Marketplace Governance.*

The corresponding conceptual model representing the implication of architectural aspects to the threat landscape of the data marketplaces is illustrated in **Figure 23**.

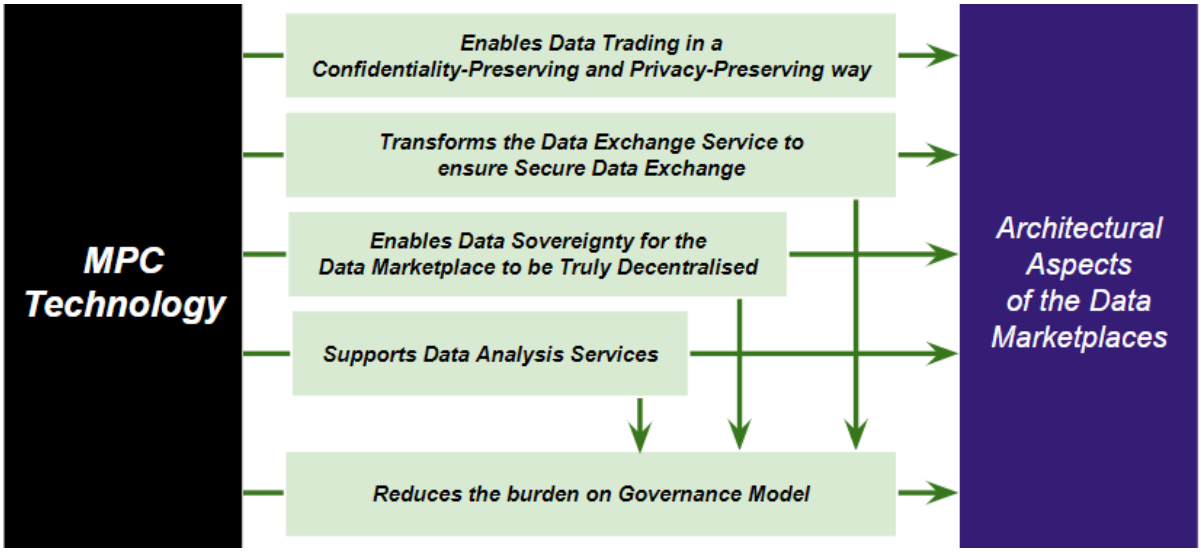

Figure 23: Architectural Implication of MPC technology to the Data Marketplaces

7.2 Implication to the Threat Landscape of Data Marketplaces

The conceptual model representing the second half of the theoretical framework is described here. The conceptual model reflects the implication of the MPC technology to the threat landscape of the data marketplaces by explicating the difference between *Pre-MPC Threat Model 2.0* and *Post-MPC Threat Model 2.0*; further generalizing the same for the business species of data marketplaces.

Affects the Business Threat Landscape both positively and negatively
MPC technology could mitigate few of the business threats associated with the sensitive nature of data. This aspect could be attractive for all the actors in the data marketplace ecosystem as it overcomes the main concerns associated with commoditizing and trading data. On the other hand, MPC could also affect negatively on some of the threats posing more threat than mitigating them. Mitigation of each business threat identified earlier is described as follows,

Mitigates the threat of Loss of Control over Data
MPC incorporation could enforce *Data Sovereignty* effectively enabling the truly decentralized data trading platform which overcomes the threat of data owners losing the control over their data as they hold their data at their site and MPC protocol provides the required knowledge from the data to the dedicated receiver with the help of its cryptographic blocks.

Reduces threat of Data Breach
MPC protocol holds either encrypted version of the data or the intermediate data in during the protocol execution which decapitates the threat of data breach on the communication channel as the breached data does not have any value. This way the threats of *Data Leakage* and *Data Leakage by Back Correlation* could become irrelevant by MPC incorporation. However, the breach could disrupt the protocol execution. However, the risk associated with this could be very less compared to the actual data breach.

Ensures no Loss of Competitive Advantage for Data Actors
As the data exchange happens in a confidentiality-preserving way via MPC Protocol; in the sense that only the result of the agreed upon computation is learnt to the receiver, there would be no risk of that receiver reverse engineering critical aspects of the owner's business processes; thus, MPC could overcome the threat for the data actors losing their competitive advantage when sharing data.

Decapitates Regulatory Threats because of Privacy-Preservation
The privacy-preserving nature of the MPC Protocol preserves the personal information which could be in the data provisioned by the data owner. This is an incentive for the data owners as the regulatory threats associated with privacy violation and data security are made irrelevant because of no physical data transfer or access.

On the flip side, MPC incorporation could present with shortcomings to the existing situation. The business threats that apply for the data marketplaces even after the incorporation of MPC technology into the business processes of the data marketplaces can be attributed as the negative implication of the MPC technology to the threat landscape of the data marketplaces.

Redefines the threat of Trust Issues
The threat of *Trust Issues* could get redefined into data actors not wanting to participate on the data marketplace platform as they could find it difficult to trust a technological setting of just MPC protocol to handle their valuable commercial data.

Intensifies the threat of Induction of Malicious Data Actor
MPC could incentivize malicious data actors who just wants to gain from the benefits of the data marketplace platform either by making relationships to gain insights, supplying faulty data for the protocol execution, gaining intelligence of other actors who might be competitors et cetera. If the number of malicious parties in the execution of protocol exceeds the number of honest parties, then the protocol would become invalid. This can again be attributed as a ramification of the threat of trust issues where in the technological setting the trust is implicit, and parties should trust the process blindly which is reasonable as the technology is solid. However, the presence of the malicious actors is not accounted and could manifest into trust issues.

Capacitates all threats if the protocol execution is compromised
This attributes to the underlying risk with the technology that if the protocol is compromised in some way, then all business threats associated with data sensitivity materialize.

These shortcomings can be addressed by the incorporation of the *MPC Process Auditing* as a function of the *Governance Model* which could audit the health of the incorporated MPC technological component. This basically proves that to enable data marketplaces, the right coordination between the technological and non-technological aspects is needed which can be related to this case as a right coordination between the ***Governance Model*** and the ***MPC powered Data Exchange Service.*** This can be expressed to represent a concept in the conceptual model: ***Enabling a Safe and Secure Data Trading with the right coordination from the Governance Model.*** The corresponding conceptual model representing the implication of MPC technology to the threat landscape of the data marketplaces is illustrated in **Figure 24**.

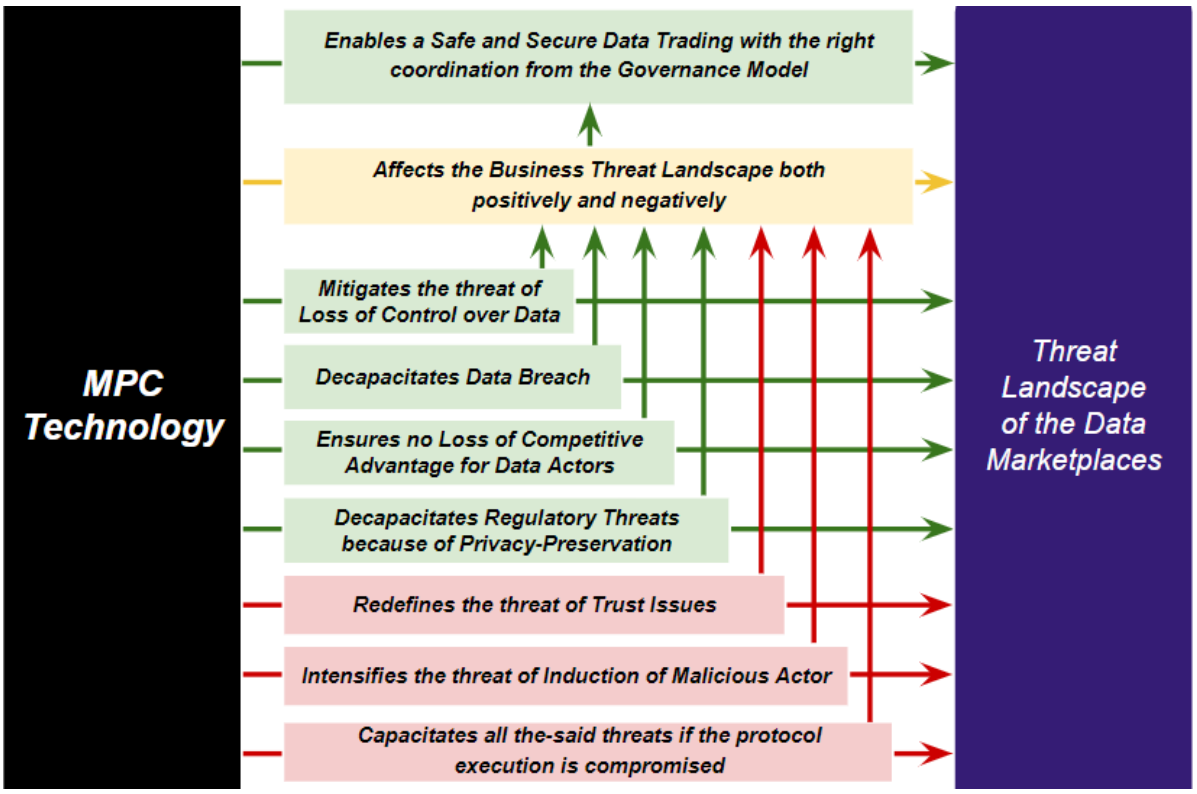

Figure 24: Implication of MPC technology to the Threat Landscape of Data Marketplaces

7.3 Incentives for the adoption of MPC technology

We find that MPC technology holds great promise for the development of efficient and secure data marketplaces. However, MPC technology has to mature to be applied in the data marketplaces. Then it could enforce safe and secure data trading in a Security-by-Design way by eliminating the business threats of ***Loss of Control over Data, Data Leakage, Data Leakage by Back Correlation, Loss of Competitive Advantage for Data Actors, Regulatory Threats*** and ***Data Sensitivity.*** MPC eliminates serious business threats associated with the issue of ***Data Sensitivity*** in a ***Security-by-Design*** way. This, in turn, reduces the burden of the ***Governance Model*** on its technological front. Hence, Incentive structures for the adoption of MPC in data marketplaces should focus on the operational benefit of a streamlined governance model. The potentially negative implications of MPC technology thereby still do not outweigh the positive synergies created by the reduced need for interparty trust and limited exposure of data assets.

8 Conclusion

This document provides the results of Task 2.1 in WP2 of Safe-DEED, threat and incentive models for distributed MPC enhanced data-marketplaces. For this task, we developed four different artifacts:

- Marketplace Architecture Model (1.0/2.0) (pre-MPC/post-MPC)
- Data Marketplace Threat Models (1.0/2.0) (pre-MPC/post-MPC threat landscape)
- A model of implications of MPC technology to the Threat Landscape of Data Marketplaces
- A list of unique selling points of MPC technology to facilitate its adoption

8.1 Summary of Threat and Architecture Models

8.1.1 Pre-MPC Data Marketplace Platform

We first obtain an architecture to reflect a generic data marketplace platform prior to incorporation of MPC technology. To this end, a literature study was conducted on data marketplaces with an aim to explore the phenomenon of data marketplaces involving their fundamental concepts like the definition, different features, relevant actors et cetera and also, to obtain an understanding of the architectural aspects of the data marketplace. However, the architectural knowledge was not found in the literature and hence, a framework was developed to build a high-level architecture for the data marketplace platform. This framework was referred as *HLA framework* and consisted of three attributes namely, Functional Requirements, Customers and Functional Components. Using the HLA framework, from the knowledge obtained through the literature study, a high-level architecture was built for a generic data marketplace assuming it to be just a technological platform. The resulting high-level architecture was subjected to validation through expert interviews and the following sub-research questions were formulated for the same.
The architecture was found to be mostly valid, while the consulted experts suggested further relevant improvements. The flagship comment was that, the data marketplace should not be viewed only as a technological platform but should be viewed as a business entity within an ecosystem. As a result, the architecture underwent significant changes to included relevant non-technological elements with respect to all the three attributes, functional requirements, actors in the ecosystem and function components, major addition being of a *Governance Model*. Furthermore, a new taxonomy for the data marketplace platform designs was developed in which our focal data marketplace was positioned as *many-to-many B2B decentralized serendipity model data marketplace.* This aspect was also incorporated along with a few more improvements and finally, an updated architecture was obtained. The final architecture of the *Pre-MPC Data Marketplace* is illustrated here in **Figure 25**.

8.1.2 Post-MPC Data Marketplace Platform

To understand the implication of the MPC technology on the high-level architecture of the data marketplace platform, we created a model for a *Post-MPC Data Marketplace Platform* by integrating *Safe-DEED Components* into the *Data Exchange Service*. As a result, the platform becomes decentralized where the actors can meet over the platform and the *Data Exchange Service* enabled by *Safe-DEED Components* is set up ad-hoc by the marketplace outside the platform, i.e., the participants execute the MPC protocol and share data. This was the direct effect of the MPC incorporation. The initial concept of MPC incorporation was subjected to the validation through expert interviews.
The conceptualization of the MPC incorporation through Safe-DEED components to gain the Post-MPC Data Marketplace Platform 1.0 was remarked as valid. However, since the pre-MPC architecture had undergone updating during the validation prior to this, the effects of the MPC incorporation were deduced for the newly obtained *Pre-MPC Data Marketplace Platform 2.0* and the changes validated and suggested by the experts were incorporated to obtain a more valid conceptualization of the incorporation of MPC in the form of *Post-MPC Data Marketplace Platform 2.0:*

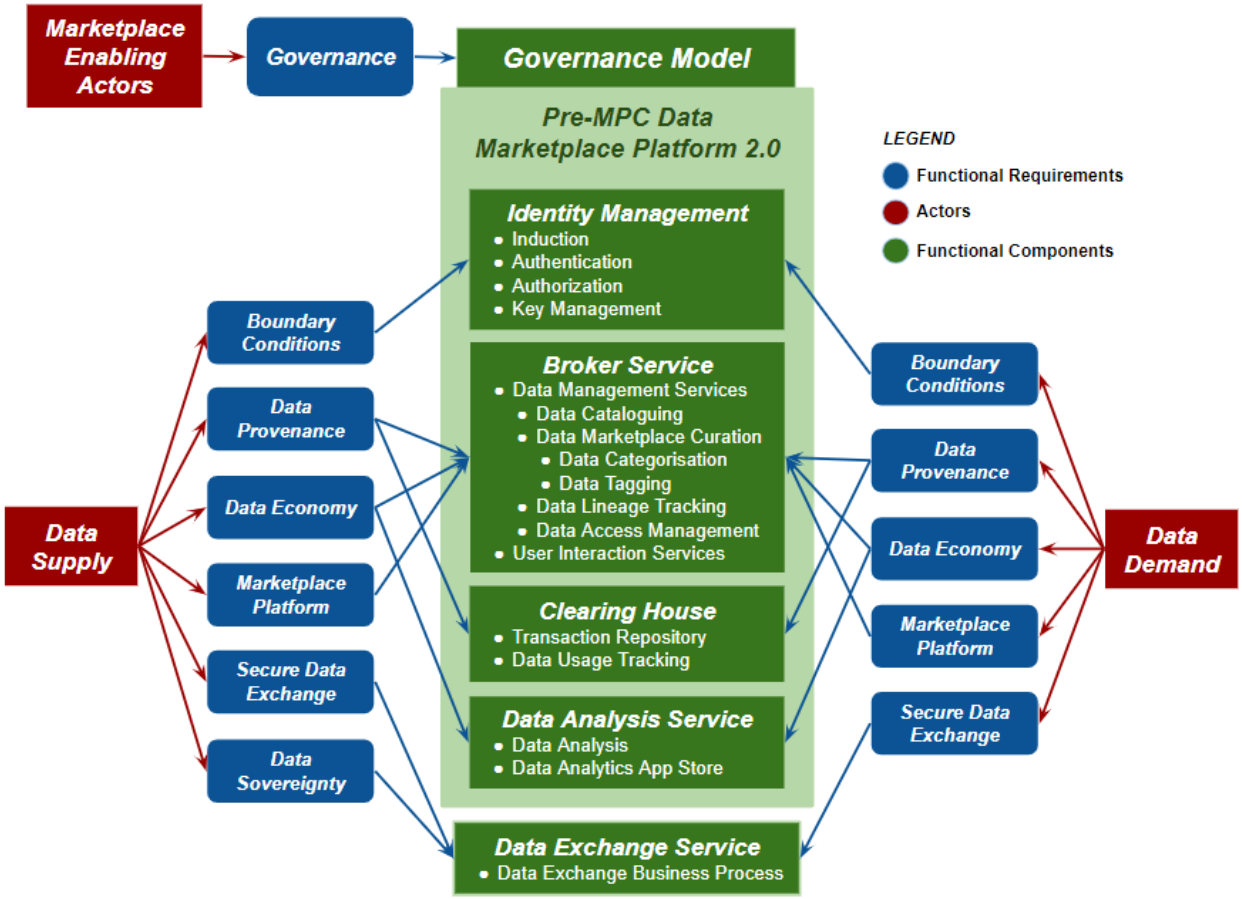

Figure 25: Pre-MPC Data Marketplace Platform 2.0

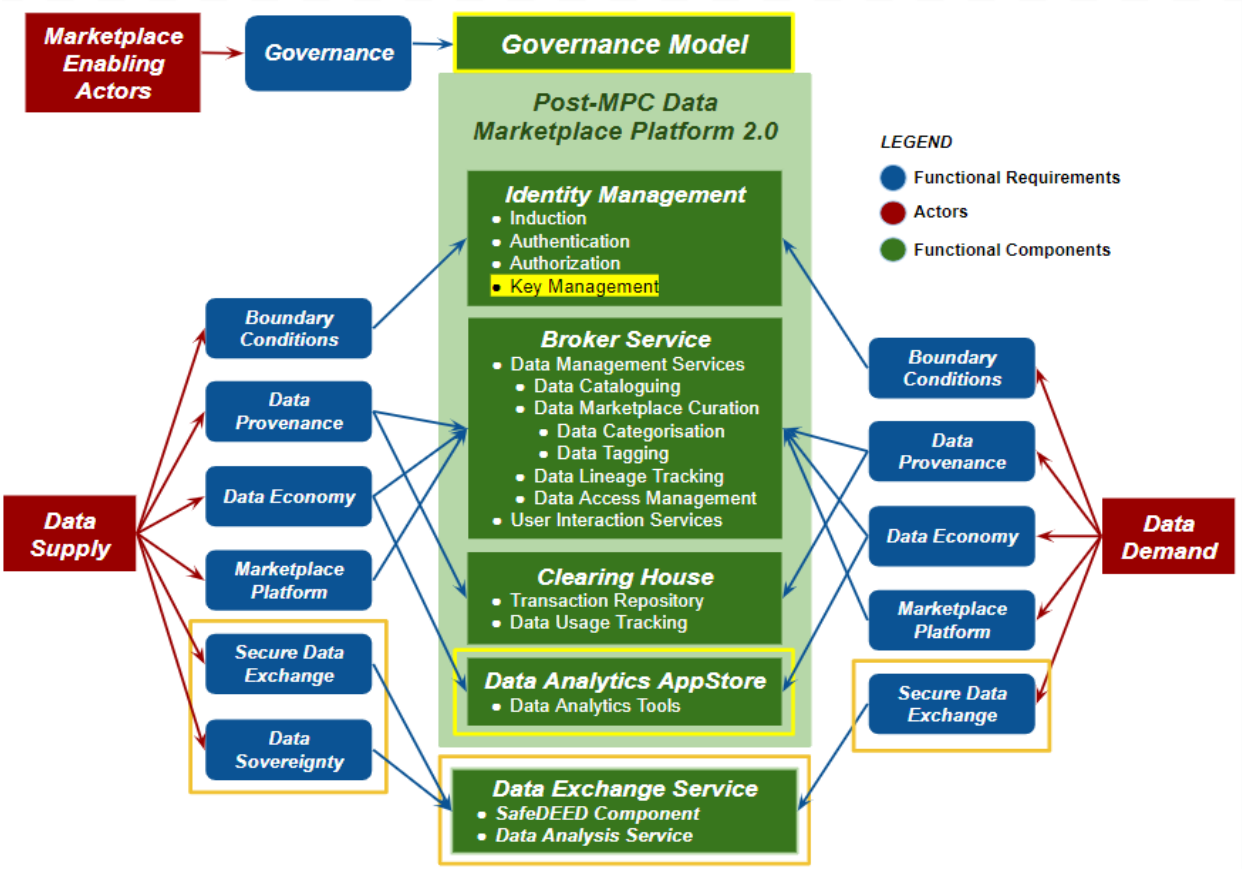

Figure 26: Post-MPC Data Marketplace Platform 2.0

- *Data Exchange Service* is provided ad-hoc with Safe-DEED component outside the platform (same conceptualization as of *Post-MPC Data Marketplace Platform 1.0*)
- Data Analysis Service is integrated into the Data Exchange Service for MPC protocols support data analysis. As a result, the platform only contains Data Analytics AppStore.
- Since the requirement of Secure Data Exchange and Data Sovereignty are enforced technologically by MPC alone, the burden is reduced from the Governance Model with respect to these requirements.
- Eliminates key management in the identity management component as it would no longer needed in the presence of MPC.

The components that received conceptual changes are highlighted in yellow/orange in **Figure 26**.

8.1.3 Pre-MPC Threat Model

To identify the threats associated with the data marketplaces, the following sub-research questions were formulated and were answered during the conceptualization phase using desk research methods. We conducted a literature study on threat modelling to understand how the process of threat modelling could be applied to the case of our high-level architecture. It was deduced that none of the threat modelling frameworks in the literature were applicable to our case of performing threat modelling at the level of business functions. Following this, the HLTM framework was developed which helps in carrying out high-level threat modelling for the technological entities with high-level architectures such as ours. Furthermore, the idea of a threat landscape was conceptualized which comprised of the combination of cyber threats and its business consequence to the focal entity. Using the HLTM framework and the literature analysis of cyber threats, a high-level threat model was developed comprising of the threats that affect each individual component in high-level architecture and the threats were coupled with their relevant business consequences to represent the threat landscape of our high-level architecture. The resulting threat model reflected the *Pre-MPC Threat Model 1.0*. Following this, the threat model was subjected to validation through expert interviews.
The *Pre-MPC Threat Model 1.0* was remarked as invalid as the threat model did not reflect the actual threat landscape of the data marketplaces. It was reasoned that the threats in the threat model consisted mostly of baseline threats as they were built upon the baseline architectural specification of individual components of the high-level architecture. Since these were basic threats which could be mitigated by incorporating appropriated mitigation techniques, the threat model was criticized as not reflecting the actual threat landscape of the data marketplaces as the threats that actually hinder data marketplaces are much more complex to address.
The experts remarked that those kinds of threats affect the business logic at a level much higher than our low component level analysis. However, these threats were named as *business threats*. Hence, the scope of our analysis was heightened from the low component-level to the high business level and accordingly, the conceptualization of the threat landscape was updated to consider threats only to business logic. Further, a new threat model comprising of the business threats to the business logic i.e. data sensitivity, data handling et cetera was built. The business threats that hinder data marketplaces are:

- *Loss of Control over Data for the Data Owner* making the data owner reluctant to participate in the data marketplace platform.
- *Trust Issues for the Data Actors* as the trust is implied intangibly but nor established with explicit measures of mechanisms which makes hard for the data actors to trust each other in a business setting.
- *Data Leakage* where the data may not be used by the concerned party as stipulated in the contract and could be leaked to other parties, or data breach because of encryption failure.
- *Data Leakage by Back Correlation:* special case where the anonymized data can be coupled with auxiliary data to obtain the personal information which was anonymized earlier. In addition to PII, even confidential information can be obtained by combining the data with appropriate auxiliary data.

- *Loss of Competitive Advantage for the Data Actors* where the actors owning data feel reluctant to share their data as they fear if that data might give away proprietary information which could result in the loss of competitive advantage.
- *Regulatory Threats* become relevant if the private information is involved and transacted without complying to the GDPR which may cause regulatory fines and legal complications for all the actors involved.
- Data Sensitivity: relates to the sensitive nature of the data and the challenges with respect to its commodification. As long as there is physical data involved in the transaction, all the above threats prevail.
- *Induction of Malicious Actors* which is a different threat other than data sensitivity where a malicious could be inducted and he can use the platform services for his benefit while deceiving other actors with invalid participation like providing invalid data, learning about metadata of other actors et cetera.

These threats constitute the business threats which reflect the threat landscape of the data marketplaces as reflected by the experts.

8.1.4 Post-MPC Threat Model

To understand the effect of MPC on the threats modelled in *Pre-MPC Threat Model*, we incorporated the introduction of MPC to form the *Post-MPC Threat Model*. We found that MPC technology overcomes the threats associated with data handling like data breach, privacy breach etc. by bringing about structural changes in the data exchange mechanism in a Security-by-Design way, while the threats for the remaining components remain unaffected. This threat model was put up for validation in expert interviews. Our respondents established that the incorporation of MPC would overcome the following threats:

- *Loss of Control over Data*
- *Data Leakage*
- *Data Leakage by Back Correlation*
- *Loss of Competitive Advantage for Data Actors*
- *Regulatory Threats*
- *Data Sensitivity*

These threats are mitigated as MPC enables data sharing to happen in a confidentiality-preserving and privacy-preserving way such that the physical data is never transferred to different parties but are accessed in the form of an MPC protocol which only delivers computational result to the dedicated receiver. As a result, the threats associated with the sensitivity of data become irrelevant as the data resides at the owner's site safely. However, the incorporation of MPC can introduce threats which could affect data marketplaces. These threats that exist even after MPC incorporation are:

- *Trust Issues*: The threat of data actors not participating in the data marketplace as the data actors find it difficult to trust a technological setting of just MPC protocol to handle their valuable commercial data
- *Induction of Malicious Data Actor*: The threat of inducting a malicious actor into the data marketplace as a legitimate costumer. Since the tangible trust is orchestrated by technology, the malicious actor could take advantage of this feature to gain crucial information by exploiting other actors. MPC intensifies the actions of malicious actors.
- *Faulty Execution of MPC Protocol*: This is a fundamental threat where all the promise of the MPC relies on it being functioning as expected. However, if the execution goes faulty or if the protocol is compromised somehow, then all the threats mitigated by MPC would become relevant.

Accordingly, we incorporated these threats in the final *Post-MPC Threat Model 2.0.*

8.2 Key Findings

We find that MPC technology eliminates the business threats of ***Loss of Control over Data, Data Leakage, Data Leakage by Back Correlation, Loss of Competitive Advantage for Data Actors, Regulatory Threats*** and ***Data Sensitivity.*** Hence, MPC eliminates serious business threats associated with the issue of ***Data Sensitivity*** in a ***Security-by-Design*** way. This, in turn, reduces the burden of the ***Governance Model*** on its technological front. Hence, Incentive structures for the adoption of MPC in data marketplaces should focus on the operational benefit of a streamlined governance model. The potentially negative implications of MPC technology thereby still not outweigh the positive synergies created by the reduced need for interparty trust and limited exposure of data assets.

Furthermore Task 2.1 of Workpackage 2 made the following contributions:

- A new taxonomy of data marketplace platform designs, which provides an updated classification comprising of the different platform designs containing both concept platforms and realized ones. The taxonomy refines the basic classification of Koutroumpis et al. (2017) and updates it with a variety of probable data marketplaces**.** This provides a new foundation to position different data marketplaces either during design or analysis.
- A new list of functional requirements was developed which furthers the conversation of the functional requirements from just being technological to also include non-technological aspects, helping to understand what is expected of data marketplaces.
- As a significant contribution to the gap in the literature involving the architectural aspects of the data marketplaces, this research presents the *High-Level Architecture* of a generic data marketplace platform**.** This architecture can act as a reference architecture for the researchers to build more sophisticated and detailed architectures for the data marketplace platforms. Additionally, the *HLA Framework 2.0* can be used by researchers to build high-level architectures for the technological entities.
- We develop a new *Business Threat Model* and a *Cyber Threat model* for a *generic* data marketplace platform. This threat models marks the first of its kind for data marketplaces.

Secondly, the task also contributes to the state of the art for *threat modelling***,** which mostly focuses on software centric threat modelling, lacking a business function focus:

- A new taxonomy for threat models was created to expand the scope of threat modelling from just the low-level cyber threats to also include high-level business threats. This taxonomy goes beyond just focusing on the cyberspace and includes the analyses of threats to the business logic of the focal entity. The NGCI Apex Classification of Cyber Threat Models by Bodeau et al. (2018) is also positioned in our taxonomy.
- A new cyber threat modelling framework which operates at the business function level of information systems of technological entities was developed which goes by the name *High-Level Cyber Threat Modelling (HLCTM)* framework. This framework provides an effective threat model for detained architectures and provides a baseline threat model for high-level technological entities. Additionally, the framework provides a straight forward way to carry out low-profile threat modelling on technological entities which can be used for auxiliary tasks of researches in bigger scopes.

Finally, the task also contributes to the gaps existing in the literature of the *MPC technology* mostly associated with its business application aspects. These are discussed as listed below,

- Our research clarified the business process of MPC technology finding that the process is dependent on the underlying use-case and hence, cannot be standardized for a platform like data marketplace. Instead, it can only be provisioned in an ad-hoc form. Furthermore, we explicated the application of this business process in a data marketplace platform. Thus, contributing an application for the gap involving the business application of MPC technology.

Furthermore, we have also investigated the effect of MPC on the threats associated with data sensitivity and data marketplaces which furthers the literature explicating the advantages and shortcomings of MPC technology.

This page has been intentionally left blank.

9 Bibliography

Al-Mohannadi, H., Mirza, Q., Namanya, A., Awan, I., Cullen, A., & Disso, J. (2016). "Cyber-Attack Modeling Analysis Techniques: An Overview." In *2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)* (pp. 69–76). Vienna. https://doi.org/10.1109/W-FiCloud.2016.29

Arrow, K. J. (1972). "Economic Welfare and the Allocation of Resources for Invention." In *Rowley C.K. (eds) Readings in Industrial Economics* (pp. 219–236). Palgrave, London. https://doi.org/10.1007/978-1-349-15486-9_13

Bishop, M. (1991). *"An Overview of Computer Viruses in a Research Environment."* Dartmouth College, Hanover, NH, USA. Retrieved from http://www.ncstrl.org:8900/ncstrl/servlet/search?formname=detail%5C&id=oai%3Ancstrlh%3Adartmouthcs%3Ancstrl.dartmouthcs%2F%2FPCS-TR91-156

Bodeau, D., & Graubart, R. (2014). "A Framework for Describing and Analyzing Cyber Strategies and Strategic Effects", MTR 140346, PR 14-3407. The MITRE Corporation, Bedford, MA.

Bodeau, D. J., Mccollum, C. D., & Fox, D. B. (2018). *"Cyber Threat Modeling: Survey, Assessment, and Representative Framework"," PR 18-1174*. HSSEDI, The MITRE Corporation. Retrieved from https://www.mitre.org/sites/default/files/publications/pr_18-1174-ngci-cyber-threat-modeling.pdf

Brynjolfsson, E., & McAffee, A. (2012). "Big Data: The Management Revolution." *Harvard Business Review*, *90*(10), 60–68. Retrieved from http://tarjomefa.com/wp-content/uploads/2017/04/6539-English-TarjomeFa-1.pdf

Chakrabarti, A., Quix, C., Geisler, S., Khromov, A., & Jarke, M. (2018). "Goal-Oriented Modelling of Relations and Dependencies in Data Marketplaces." In *iSTAR@CAiSE 2018*. Retrieved from https://pdfs.semanticscholar.org/29ce/33d36953534defc34dcf7b01f14a7a02d0c2.pdf?_ga=2.67273746.246347384.1566747501-1794706104.1555951664

Conti, M., Dragoni, N., & Lesyk, V. (2016). "A Survey of Man In The Middle Attacks." In *IEEE Communications Surveys & Tutorials* (Vol. 18, pp. 2027–2051). https://doi.org/10.1109/COMST.2016.2548426

Corbin, J., & Strauss, A. (1990). "Grounded Theory Research: Procedures, Canons, and Evaluative Criteria." *Qualitative Sociology*, *13*(1), 3–21. https://doi.org/https://doi.org/10.1007/BF00988593

Davenport, T. H. (2006, January). "Competing on analytics." *Harvard Business Review*, *84*(1), 98–107. Retrieved from https://hbr.org/2006/01/competing-on-analytics

de Reuver, M. (2019a). MOT2312 Research Methods - 3.1. "Data Collection Operationalization." Faculty of TPM, TU Delft, Delft.

de Reuver, M. (2019b). MOT2312 Research Methods - 7.2 "Qualitative Data Analysis." Faculty of TPM, TU Delft, Delft.

Deichmann, J., Heineke, K., Reinbacher, T., & Wee, D. (2016). *"Creating a successful Internet of Things data marketplace." McKinsey & Company*. Retrieved from https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/creating-a-successful-internet-of-things-data-marketplace

Dhillon, D. (2011). "Developer-driven threat modeling: Lessons learned in the trenches." In *IEEE Security and Privacy* (Vol. 9, pp. 41–47). https://doi.org/10.1109/MSP.2011.47

Eisenmann, T. R., Parker, G., & Van Alstyne, M. W. (2006). "Strategies for Two-Sided Markets." *Harvard Business Review*, *84*(10), 92–101. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2409276

Federal Financial Institutions Examination Council. (2016). *"FFIEC Information Technology Examination Handbook: Information Security."* Retrieved from https://ithandbook.ffiec.gov/ITBooklets/FFIEC_ITBooklet_InformationSecurity.pdf

Fricker, S. A., & Maksimov, Y. V. (2017). "Pricing of data products in data marketplaces." In *Ojala A., Holmström Olsson H., Werder K. (eds) Software Business. ICSOB 2017. Lecture Notes in Business Information Processing* (Vol. 304, pp. 49–66). Springer, Cham. https://doi.org/https://doi.org/10.1007/978-3-319-69191-6_4

Fu, X. (2005). *"On traffic analysis attacks and countermeasures", Doctoral Dissertation*. Texas A&M University. Retrieved from http://hdl.handle.net/1969.1/4968

Goldreich, O. (1998). *"Secure Multi-Party Computation", Manuscript, Preliminary Version*. Retrieved from https://www.researchgate.net/publication/2934115

Guszcza, J., Steier, D., Lucker, J., Gopalkrishnan, V., & Lewis, H. (2013). *"Big Data 2.0: New business strategies from big data." Deloitte Review*. Retrieved from https://www2.deloitte.com/insights/us/en/deloitte-review/issue-12/big-data-2-0.html

Hansman, S., & Hunt, R. (2005). "A taxonomy of network and computer attacks." *Computers & Security*, *24*(1), 31–43. https://doi.org/10.1016/J.COSE.2004.06.011

Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2016). "Capturing value from big data – a taxonomy of data-driven business models used by start-up firms." *International Journal of Operations & Production Management*, *36*(10), 1382–1406. https://doi.org/10.1108/IJOPM-02-2014-0098

Hynes, N., Dao, D., Yan, D., Cheng, R., & Song, D. (2018). "A demonstration of sterling: a privacy-preserving data marketplace." In *Proceedings of the VLDB Endowment* (Vol. 11, pp. 2086–2089). https://doi.org/10.14778/3229863.3236266

Johnson, C. S., Badger, M. L., Waltermire, D. A., Snyder, J., & Skorupka, C. (2016). *"Guide to Cyber Threat Information Sharing", NIST Special Publication 800-150*. National Institute of Standards and Technology, U.S. Department of Commerce. https://doi.org/10.6028/NIST.SP.800-150

Jones, J. A. (2005). *"An Introduction to Factor Analysis of Information Risk (FAIR)"* (Vol. 1). Risk Management Insight. Retrieved from http://riskmanagementinsight.com/media/documents/FAIR_Introduction.pdf

Kamatchi, R., & Ambekar, K. (2016). "Analyzing Impacts of Cloud Computing Threats in Attack based Classification Models." *Indian Journal of Science and Technology*, *9*(21). https://doi.org/10.17485/ijst/2016/v9i21/95282

Koutroumpis, P., & Leiponen, A. (2013). "Understanding the value of (big) data." In *2013 IEEE International Conference on Big Data, Big Data 2013* (pp. 38–42). Silicon Valley, CA. https://doi.org/10.1109/BigData.2013.6691691

Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2017). *"The (Unfulfilled) Potential of Data Marketplaces." ETLA Working Papers* (Vol. 2420). The Research Institute of the Finnish Economy. Retrieved from http://hdl.handle.net/10419/201268

Kraus, L., Fiebig, T., Miruchna, V., Moller, S., & Shabtai, A. (2015). "Analyzing End-users'' Knowledge and Feelings Surrounding Smartphone Security and Privacy"." In *IEEE Security & Privacy Workshops - Mobile Security Technologies (MoST)*. San Jose, CA. Retrieved from http://www.ieee-security.org/TC/SPW2015/MoST/papers/s1p2.pdf

Leiponen, A., & Thomas, L. D. W. (2016). "Big data commercialization." In *IEEE Engineering Management Review* (Vol. 44, pp. 74–90). https://doi.org/10.1109/EMR.2016.2568798

Liang, F., Yu, W., An, D., Yang, Q., Fu, X., & Zhao, W. (2018). "A Survey on Big Data Market: Pricing, Trading and Protection." In *IEEE Access* (Vol. 6, pp. 15132–15154). https://doi.org/10.1109/ACCESS.2018.2806881

Lupu, M. (2018). *"Safe-DEED: Safe Data Enabled Economic Development" Project Proposal*. KNOW-CENTER GMBH RESEARCH CENTER FOR DATA-DRIVEN BUSINESS & BIG DATA ANALYTICS, Austria.

Marback, A., Do, H., He, K., Kondamarri, S., & Xu, D. (2013). "A threat model-based approach to security testing." *Software - Practice and Experience*, *43*(2), 241–258. https://doi.org/10.1002/spe.2111

Marotta, A., Carrozza, G., Battaglia, L., Montefusco, P., & Manetti, V. (2013). "Applying the SecRAM methodology in a CLOUD-based ATM environment." In *2013 International Conference on Availability, Reliability and Security,* (pp. 807–813). Regensburg. https://doi.org/10.1109/ARES.2013.108

Meier, J. D., Mackman, A., Dunner, M., Vasireddy, S., Escamilla, R., & Murukan, A. (2003). *"Improving web application security: Threats and Countermeasures", Satyam Computer Services, Microsoft Corporation*. Retrieved from https://docs.microsoft.com/en-us/previous-versions/msp-n-p/ff649874(v%3Dpandp.10)

Muckin, M., & Fitch, S. C. (2017). "A Threat-Driven Approach to Cyber Security: Methodologies, Practices and Tools to Enable a Functionally Integrated Cyber Security Organization." *Lockheed Martin Corporation*, 1–45. Retrieved from http://ce.sharif.edu/courses/95-96/2/ce746-1/resources/root/Resources/Lockheed Martin Threat-Driven Approach whitepaper.pdf

Muscat, I. (2019). "What Are Injection Attacks." *Acunetix*. Retrieved from https://www.acunetix.com/blog/articles/injection-attacks/

Muschalle, A., Stahl, F., Löser, A., & Vossen, G. (2013). "Pricing approaches for data markets." *Castellanos M., Dayal U., Rundensteiner E.A. (Eds) Enabling Real-Time Business Intelligence. BIRTE 2012. Lecture Notes in Business Information Processing. Springer, Berlin, Heidelberg*, *154*, 129–144. https://doi.org/10.1007/978-3-642-39872-8_10

NIST. (2011). *"Managing Information Security Risk: Organization, Mission, and Information System View", NIST Special Publication 800-39* (Vol. 40). National Institute of Standards and Technology, U.S.Department of Commerce, Gaithersburg. https://doi.org/10.1108/k.2011.06740caa.012

NIST. (2012). *"Guide for Conducting Risk Assessments", NIST Special Publication 800-30 Revision 1*. National Institute of Standards and Technology, U.S. Department of Commerce, Gaithersburg. https://doi.org/10.6028/NIST.SP.800-30r1

Northcutt, S. (2018). *"Security Controls."* Retrieved from https://www.sans.edu/cyber-research/security-laboratory/article/security-controls

Omotunde, H., & Ibrahim, R. (2015). "A review of threat modelling and its hybrid approaches to software security testing." *ARPN Journal of Engineering and Applied Sciences*, *10*(23), 17657–17664. Retrieved from http://www.arpnjournals.org/jeas/research_papers/rp_2015/jeas_1215_3222.pdf

Osterwalder, A., Pigneur, Y., Bernarda, G., & Smith, A. (Designer). (2014). Value proposition design : how to create products and services customers want. Retrieved from https://www.wiley.com/en-us/Value+Proposition+Design%3A+How+to+Create+Products+and+Services+Customers+Want-p-9781118968055

OWASP. (2018). "Forced browsing", Open Web Application Security Project,. Retrieved July 1, 2019, from https://www.owasp.org/index.php/Forced_browsing

Parker, D. B. (2015). "Toward a New Framework for Information Security?" In *Computer Security Handbook (eds S. Bosworth, M. E. Kabay and E. Whyne)* (pp. 3.1-3.23). Hoboken, NJ, USA. https://doi.org/10.1002/9781118851678.ch3

Quix, C., Chakrabarti, A., Kleff, S., & Pullmann, J. (2017). "Business process modelling for a Data Exchange Platform." In *Proc. Forum at the 29th International Conference on Advanced*

Information Systems Engineering (CAiSE), CEUR Workshop Proceedings (Vol. 1848, pp. 153–160). Essen, Germany, 2017. Retrieved from http://ceur-ws.org/Vol-2118/iStar2018_paper_4.pdf

Ramel, D. (2016). "Microsoft Closing Azure DataMarket." *Application Development Trends MAG*. Retrieved from https://adtmag.com/articles/2016/11/18/azure-datamarket-shutdown.aspx

Roman, D., & Stefano, G. (2016). "Towards a Reference Architecture for Trusted Data Marketplaces: The Credit Scoring Perspective." In *2016 2nd International Conference on Open and Big Data (OBD)* (pp. 95–101). Vienna. https://doi.org/10.1109/OBD.2016.21

Rosenquist, M. (2009). *"Prioritizing Information Security Risk with Threat Agent Risk Assessment." IT@Intel White Paper*. Retrieved from http://media10.connectedsocialmedia.com/intel/10/5725/Intel_IT_Business_Value_Prioritizing_Info_Security_Risks_with_TARA.pdf

Sabbagh, B. Al, & Kowalski, S. (2015). "A Socio-technical Framework for Threat Modeling a Software Supply Chain." *IEEE Security and Privacy*, *13*(4), 30–39. https://doi.org/10.1109/MSP.2015.72

Schomm, F., Stahl, F., & Vossen, G. (2013). "Marketplaces for Data: An Initial Survey." *SIGMOD Rec.*, *42*(1), 15–26. https://doi.org/10.1145/2481528.2481532

Sekaran, U., & Bougie, R. (2013). *"Research Methods for Business: A Skill-Building Approach." Wiley* (Seven). Retrieved from https://www.wiley.com/en-nl/Research+Methods+For+Business:+A+Skill+Building+Approach,+7th+Edition-p-9781119266846

Simonite, T. (2016). "Technical Roadblock Might Shatter Bitcoin Dreams", MIT Technology Review. Retrieved from https://www.technologyreview.com/s/600781/technical-roadblock-might-shatter-bitcoin-dreams/

Smith, D. A. (2017). "7 Steps of a Cyber Attack and What You Can Do to Protect Your Windows Privileged Accounts", Beyond Trust. Retrieved from https://www.beyondtrust.com/blog/entry/7-steps-cyber-attack-can-protect-windows-privileged-accounts

Smith, G., Ofe, H. A., & Sandberg, J. (2016). "Digital Service Innovation from Open Data: Exploring the Value Proposition of an Open Data Marketplace." In *49th Hawaii International Conference on System Sciences (HICSS)* (pp. 1277–1286). Koloa, HI. https://doi.org/10.1109/HICSS.2016.162

Smith, J. (2018). "Data Marketplaces: The Holy Grail of our Information Age", Towards Data Science, Medium. Retrieved from https://towardsdatascience.com/data-marketplaces-the-holy-grail-of-our-information-age-403ef569fffb

Spiekermann, M., Tebernum, D., Wenzel, S., & Otto, B. (2018). "A metadata model for data goods." In *Multikonferenz Wirtschaftsinformatik (MKWI)* (pp. 326–337). Retrieved from http://mkwi2018.leuphana.de/wp-content/uploads/MKWI_147.pdf

Stahl, F., Schomm, F., Vomfell, L., & Vossen, G. (2015). *"Marketplaces for Digital Data: Quo Vadis?" Working Papers, ERCIS - European Research Center for Information Systems 24* (Vol. 24). Retrieved from http://hdl.handle.net/10419/129780

Stahl, F., Schomm, F., & Vossen, G. (2014). "Data marketplaces: An emerging species." *Frontiers in Artificial Intelligence and Applications*, *Databases*(August 2013), 145–158. https://doi.org/10.3233/978-1-61499-458-9-145

Stahl, F., Schomm, F., Vossen, G., & Vomfell, L. (2016). "A classification framework for data marketplaces." *Vietnam Journal of Computer Science*, *3*(3), 137–143. https://doi.org/10.1007/s40595-016-0064-2

Steven, J. (2010). "Threat Modeling-Perhaps It's Time." *IEEE Security and Privacy*, *8*(3), 83–86. https://doi.org/10.1109/MSP.2010.110

The MITRE Corporation. (2015). “Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK).” Retrieved from https://attack.mitre.org/

Uzunov, A. V., & Fernandez, E. B. (2014). An extensible pattern-based library and taxonomy of security threats for distributed systems. *Computer Standards and Interfaces*, *36*(4), 734–747. https://doi.org/10.1016/j.csi.2013.12.008

van Bommel, P., van Gils, B., Proper, H. A., van Vliet, M., & van der Weide, T. P. (2005). “The Information Market: Its Basic Concepts and Its Challenges.” In *Ngu A.H.H., Kitsuregawa M., Neuhold E.J., Chung JY., Sheng Q.Z. (eds) Web Information Systems Engineering – WISE 2005. WISE 2005. Lecture Notes in Computer Science* (pp. 577–583). Springer, Berlin, Heidelberg. https://doi.org/10.1007/11581062_50

Watson, C., & Zaw, T. (2018). *“OWASP Automated Threat Handbook: Web Applications”, Open Web Application Security Project*. Retrieved from https://www.owasp.org/images/3/33/Automated-threat-handbook.pdf

Wells, D. (2017). *“The Rise of the Data Marketplace: Data as a Service.”* Retrieved from https://www.datawatch.com/wp-content/uploads/2017/03/The-Rise-of-the-Data-Marketplace.pdf

Wynn, J. (2014). “Threat Assessment and Remediation Analysis (TARA).” *The MITRE Corporation*, 14–2359. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/1016629.pdf

Xiong, W., & Lagerström, R. (2019). “Threat modeling – A systematic literature review.” *Computers and Security*, *84*, 53–69. https://doi.org/10.1016/j.cose.2019.03.010

Yao, A. C. (1982). “Protocols for Secure Computations.” In *23rd Annual Symposium on Foundations of Computer Science (SFCS ’82). IEEE Computer Society* (pp. 160–164). Washington, DC, USA. https://doi.org/10.1109/SFCS.1982.88

Zhang, L., Yu, S., Wu, D., & Watters, P. (2011). “A Survey on Latest Botnet Attack and Defense.” In *2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications* (pp. 53–60). Changsha: IEEE. https://doi.org/10.1109/TrustCom.2011.11

Zlomislic, V., Fertalj, K., & Sruk, V. (2014). “Denial of service attacks: An overview.” In *2014 9th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1–6). Barcelona: IEEE. https://doi.org/10.1109/CISTI.2014.6876979

Zyskind, G., Nathan, O., & Pentland, A. (2015). “Enigma: Decentralized Computation Platform with Guaranteed Privacy.” *Cryptography and Security, Cornell University*, *abs/1506.0*, 1–14. https://doi.org/978-1-61208-153-3