Project title: Multi-Owner data Sharing for Analytics and Integration respecting Confidentiality and OWNer control
Project acronym: MOSAICrOWN
Funding scheme: H2020-ICT-2018-2
Topic: ICT-13-2018-2019
Project duration: January 2019 – December 2021

D2.5
Final Evaluation Report from Use Cases

Editors: Pierre-Antoine Champin (W3C)
Reviewers: Stefano Paraboschi (UNIBG)
Pierangela Samarati (UNIMI)

Abstract

In this deliverable, we present how the prototypes (described in Deliverable D2.4) address the requirements that were identified in Deliverable D2.1. For each of the use cases considered in MOSAICrOWN, we recall the list of functional and non-functional requirements associated with it, give a synthetic overview of the innovation developed in the course of the project, and assess how each requirement was fulfilled.

Type	Identifier	Dissemination	Date
Deliverable	D2.5	Public	2021.12.31

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825333.

MOSAICrOWN Consortium

1.	Università degli Studi di Milano	UNIMI	Italy
2.	EMC Information Systems International	EISI	Ireland
3.	Mastercard Europe	MC	Belgium
4.	SAP SE	SAP SE	Germany
5.	Università degli Studi di Bergamo	UNIBG	Italy
6.	GEIE ERCIM (Host of the W3C)	W3C	France

Versions

Version	Date	Description
0.1	2021.11.29	Initial Release
0.2	2021.12.20	Second Release
1.0	2021.12.31	Final Release

List of Contributors

This document contains contributions from different MOSAICrOWN partners. Contributors for the chapters of this deliverable are presented in the following table.

Chapter	Author(s)
Executive Summary	Pierre-Antoine Champin (W3C)
Chapter 1: Introduction	Pierre-Antoine Champin (W3C), Rigo Wenning (W3C)
Chapter 2 : Use Case 1 (EISI)	Aidan O Mahony (EISI)
Chapter 3 : Use Case 2 (MC)	Flora Giusto (MC), Saverio Mucci (MC)
Chapter 4 : Use Case 3 (SAP SE)	Jonas Böhler (SAP SE)
Chapter 5 : Conclusions	Pierre-Antoine Champin (W3C)

Contents

List of Figures

List of Tables

Executive Summary

The partners of the MOSAICrOWN project have collectively built a suite of tools and methods for unlocking the potential of data markets. By allowing data to be wrapped, sanitized, and subjected to policies, they make it possible for multiple parties to share and analyze data in a way that complies with regulations (like GDPR), respects users' privacy, and protects each party's business confidential data.

The MOSAICrOWN approach was implemented in three different use cases, each lead by one industrial partner, applied to different fields of business, and focusing on different parts of the approach. Use Case 1, led by EISI, considers data produced by Intelligent Connected Vehicles and by charging station infrastructures. It explores all protection techniques, focusing on the ingestion phase. Use Case 2, led by MC, considers transaction-level financial data. It focuses on wrapping (reversible) protection techniques and on policies, at different stages of the data lifecycle. It also focuses and the compliance with different regulations. Use Case 3, led by SAP SE, considers consumer analytics cloud-based data markets where storage and/or analysis can be distributed. It focuses on sanitization (non-reversible) protection techniques, and the privacy-utility trade-offs they offer. It addresses all stages of the data lifecycle. The prototypes developed in each of these use cases are described in Deliverable D2.4 "Use Case Prototypes".

The functional and non-functional requirements of each use case have been analyzed and listed in Deliverable D2.1 "Requirements from the Use Cases". They have then been continuously monitored during the project, and their progress was reported in D2.2 "Report on Requirements, Research Alignment and Deployment Plan" and D2.3 "Final Report on Research Alignment". In this deliverable, we present the innovations developed in each use case in the course of the project, and assess how they address each identified requirement.

1. Introduction

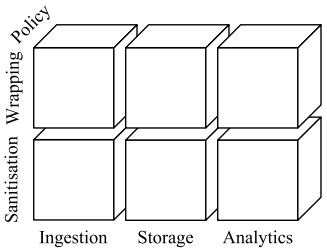
With its policy furthering the creation of a data economy, the European Commission intends to unleash the huge potential of the enormous amounts of data created in our digital society every day [Com20d]. Exploiting data at this scale creates risks for society. The European legislator introduced a number of rules to mitigate those risks, like GDPR [Uni16] and continues to do so with proposals for a Digital Market Act [Com20b], a Digital Services Act [Com20c], a Data Governance Act [Com20a], and proposed legislation on Artificial Intelligence [Com21].

Those rules, while mitigating risks, create new hurdles for data processing, and thus also slow the exploitation of the full potential of the data produced in our digital society. Consequently, the challenge is to create responsible data processing tools that permit to exploit a high percentage of the potential of the data while mitigating the risks in a balanced way. This challenge is not simple to resolve. MOSAICrOWN took it up and promised to create tools to help implementing this balanced approach between data exploitation and risk mitigation. At the same time, it was important to provide tools for computer-aided compliance to the many old and new rules governing data processing.

On a conceptual level, MOSAICrOWN started with work on functionally and ethically-oriented requirements. This document evaluates where the requirements met difficulties and where the concept underlying MOSAICrOWN worked out and allowed for balanced processing and increased, but responsible, data sharing. Deliverable D2.1 “Requirements from the Use Cases” was the first document created by MOSAICrOWN. The requirements were designed in line with the MOSAICrOWN ethical guidelines delivered as D1.1 “POPD - Requirement No. 2”.

MOSAICrOWN relies on three pillars to engineer the balanced processing and compliance:

1. **Policy** to attach rights and use limitations to data in order to manage the processing of data according to the rules set by the environment, users and other policies.
2. **Wrapping** to provide tools to obfuscate and secure data at certain points of the date lifecycle, including at ingestion time and when serving in the data market, e.g. by encrypting the data.
3. **Sanitization** to clean data at various points of the data lifecycle, e.g. before sharing or before doing analytics or machine learning. The most prominent example is anonymization.



All three methods can be combined or used in isolation. MOSAICrOWN has three uses cases (UC). Everyone of them had a specific focus on a technique. UC1 on Intelligent Connected Vehicles (ICV) focused on the combined use of policy, wrapping and sanitization, during the ingestion phase. UC2 and UC3 have their focus in wrapping and sanitization, respectively.

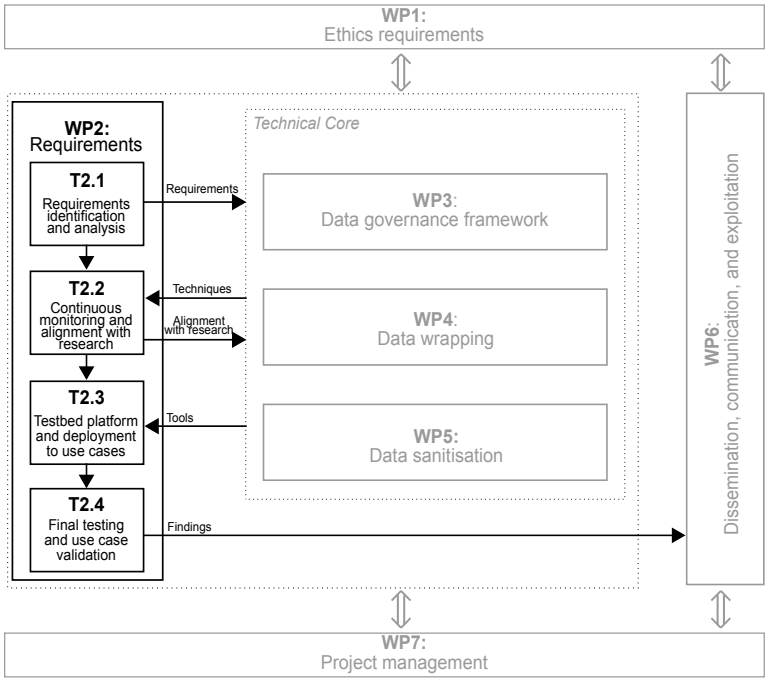


Figure 1.1: Interactions between WP2 and other WPs within MOSAICrOWN

Despite the fact that certain use cases had a certain focus, partners of the MOSAICrOWN project have collectively built a suite of tools and methods for unlocking the potential of data markets falling into the three aforementioned categories. Great care was taken to ensure combinability and complementarity between the methods. A policy may state which wrapping technique to use in one of the transforming steps to which data are submitted before reaching the data market. Cloud storage used within the framework of a data fabric may require cryptographic means as used in the wrapping technique. And we all love statistics and machine learning. UC3 yielded in tools not only being able to sanitize and anonymize data with very modern and robust methods. Before doing so, we also evaluated the risks attached and the level of anonymity required and the tool could react on this type of semantics expressed in policy. Additionally, the presence of policy annotations in the data fabric removes uncertainty about permissions to process data. Removing the high liability risk, the data lake is transformed from toxic digital waste into an intelligent data fabric.

MOSAICrOWN followed an agile approach. We started with the requirements, but those were not strictly carved in stone. With feedback from implementation and research MOSAICrOWN had the option to adapt certain criteria. The goal of Work Package 2 (WP2) is to coordinate the use cases considered in the project, provide requirements, deployment and validation of MOSAICrOWN solutions, and enable direct exploitation by the industrial partners (see Figure 1.1 for a synthetic representation of the interactions of WP2 with the other WPs of MOSAICrOWN).

This deliverable concludes the work in this agile approach and evaluates in how far requirements have been met or had to be adapted. Adaptation does not necessarily mean to change a requirement or to add a requirement. It was also possible to give existing requirements an interpretation that was not present during the writing of D2.1 “Requirements from the Use Cases”. Finally, the goal is to assess the prototypes developed for each use case (described in Deliverable D2.4 “Use Case Prototypes”) in that respect.

The remainder of the deliverable is organized as follows. Chapter 2 describes Use Case 1, led

by EISI, which considers data produced by Intelligent Connected Vehicles and by charging station infrastructures. It explores all protection techniques, focusing on the ingestion phase. Chapter 3 is about Use Case 2, led by MC. This use case considers transaction-level financial data. It focuses on wrapping (reversible) protection techniques and policies at different stage of the data lifecycle, and the compliance with different regulations. Chapter 4 presents the results of Use Case 3, led by SAP SE, and considering consumer analytics cloud-based data markets where storage and/or analysis can be distributed. It focuses on sanitization (non-reversible) protection techniques, and the privacy-utility trade-offs they offer. Finally, in Chapter 5, we conclude with a brief synthesis of the findings of the previous chapters.

2. Use Case 1 (EISI)

This chapter reports on the final evaluation of Use Case 1 (UC1) with regards to how the tools developed during MOSAICrOWN align with the requirements defined in Deliverable D2.1 "Requirements from the Use Cases".

To provide context we briefly remind the reader as to the goals and motivation of UC1. This use case involves multiple parties bringing together disparate data sources in the context of Intelligent Connected Vehicles (ICV) for the provision of monetized services that will create new data markets in an emerging sector. The scenario proposed in UC1 demonstrates the effectiveness of MOSAICrOWN data protection mechanisms in real situations where a viable ICV data market requires the application of a data governance model and supporting data wrapping and sanitization, to provide secure, granular access to data sharing and analytics. The actors involved in the use case are: the connected vehicle fleet, the Electric Vehicle (EV) charging infrastructure provider, and the MOSAICrOWN cloud provider hosting the data sharing and analytics platform (Figure 2.1).

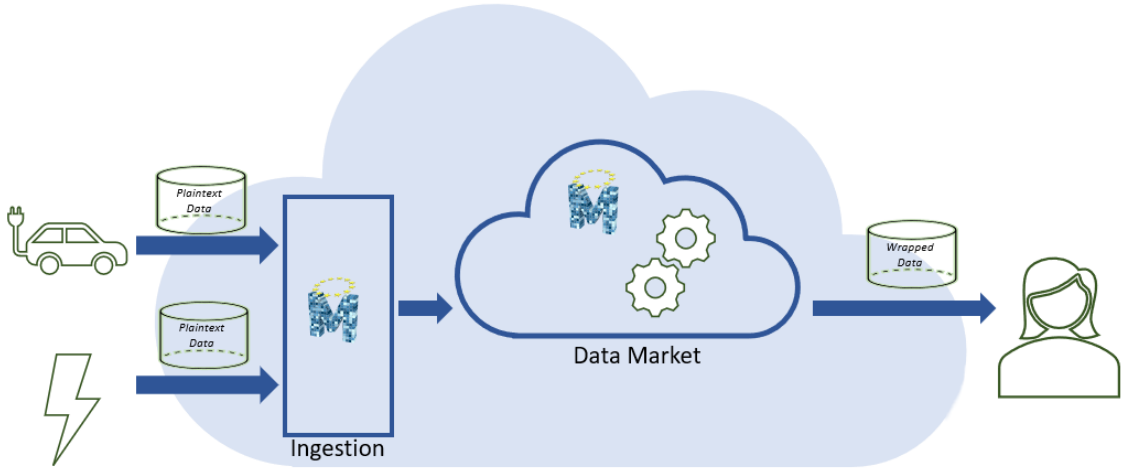


Figure 2.1: Flexible data sharing: plaintext data, sanitized data, or wrapped data suitable for secure computation

The goal of UC1 is to facilitate the connected vehicle fleet manager and the EV charging infrastructure provider exchanging data such that they can derive mutually beneficial insights into the status of the EV charging infrastructure. UC1 is concerned primarily with the ingestion phase of the data lifecycle as visualized in Figure 2.2.

In the remainder of this chapter, we first detail the requirements considered for UC1 in Section 2.1. In Section 2.2, we detail the innovations produced during MOSAICrOWN to realize the use case. In Section 2.3, we evaluate how the developed techniques and tools address the requirements and assess their fit for practical applications.

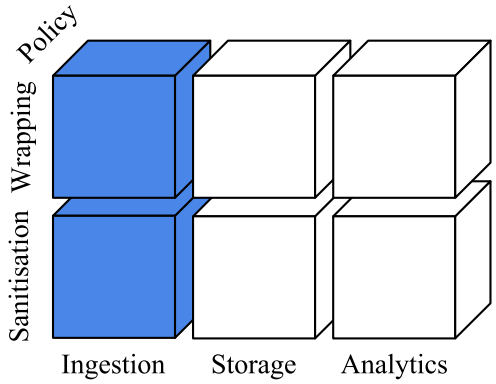


Figure 2.2: Dimensions covered by Use Case 1

2.1 Requirements

In this section, we overview the requirements to realize UC1 as defined in Deliverable D2.1 "Requirements from the Use Cases".

Table 2.1 lists the functional requirements with a short description and the dimension covered by the requirement. The functional requirements can be grouped in five aspects as summarized next.

Data Ingestion (DI) concerns itself with the facilitating the transfer of data from the data sources, i.e., the ICV and the EV charging infrastructure, into the data market.

Data Governance (DG) is required in order to allow control of the data by the owner of the data by providing a language for supporting definition and enforcement of access control rules.

Access Control Management (AC) controls the access and authorization mechanisms surrounding the granting and revoking access of data and also the sharing of that data.

Data Management (DM) facilitates the tracking of data movement as well as various guarantees over data removal and data protection.

Data Processing (DP) controls the analytics and outputs of those analytic processes within the data market.

Requirement Reference	Description	Dimension
REQ-UC1-DI1	Close to source deployment	Ingestion
REQ-UC1-DI2	Real-time stream handling	Ingestion
REQ-UC1-DI3	Batch handling	Ingestion
REQ-UC1-DI4	Support for different data types, structured and unstructured data	Ingestion
REQ-UC1-DI5	Data wrapping and data sanitization	Ingestion, Policies
REQ-UC1-DI6	Compression and Encryption	Ingestion
REQ-UC1-DI7	Secure identifier preservation	Ingestion, Storage, Policies
REQ-UC1-DI8	Assessment of data completeness	Ingestion
REQ-UC1-DI9	Protection from linkage attacks	Ingestion, Sanitization

REQ-UC1-DI10	Support ingestion from multiple concurrent sources	Ingestion
REQ-UC1-DG1	Language and definitions for data governance	Ingestion, Storage, Analytics, Policies
REQ-UC1-DG2	Support data governance models per data set per data provider	Ingestion, Storage, Analytics, Policies
REQ-UC1-DG3	Wrapping and sanitization parameters configurable by data owner	Ingestion, Policies
REQ-UC1-AC1	Access control and authorization	Storage
REQ-UC1-AC2	Grant and revoke access	Storage
REQ-UC1-AC3	Centralized key management infrastructure	Ingestion, Storage, Analytics, Policies
REQ-UC1-AC4	Support limitation to data sharing or analytics	Storage, Analytics, Policies
REQ-UC1-AC5	Allow data sharing between providers and consumers	Storage, Policies
REQ-UC1-AC6	Allow data sharing between multiple parties	Storage, Policies
REQ-UC1-AC7	Policies configurable by data provider	Storage, Policies
REQ-UC1-DM1	Tracking data movement and access	Ingestion, Storage, Analytics
REQ-UC1-DM2	Accessibility of data	Storage
REQ-UC1-DM3	Integrity of original data	Ingestion, Storage
REQ-UC1-DM4	Support for deletion guarantees	Storage
REQ-UC1-DM5	Protection at rest and in transfer	Ingestion, Storage, Analytics
REQ-UC1-DP1	Support storing data analytics results	Analytics, Policies
REQ-UC1-DP2	Anonymization of data analytics results	Analytics, Sanitization, Policies
REQ-UC1-DP3	Merge data analytics results with shared data	Storage, Analytics, Policies

Table 2.1: Use Case 1 functional requirements

Table 2.2 lists the non-functional requirements with a short description and the dimension covered by the requirement. The non-functional requirements given in the table cover three aspects as detailed in the following.

Data Economy (DE) develops mechanisms for monetizing the platform.

Performance (P) ensures the utility of the platform is sufficient for purpose.

Code Quality (CQ) provides for maintenance and testing of the software developed for the platform.

Requirement Reference	Description	Dimension
REQ-UC1-DE1	License model	Ingestion, Storage, Analytics
REQ-UC1-DE2	Distinction between data sharing and analytics	Storage, Analytics, Policies
REQ-UC1-P1	Limiting latency caused by wrapping and sanitization	Ingestion, Analytics, Sanitization, Wrapping
REQ-UC1-P2	Ensure utility under sanitization	Analytics, Sanitization, Policies
REQ-UC1-CQ1	Consider recommended coding practices	Ingestion, Storage, Analytics, Policies
REQ-UC1-CQ2	Code covering for testing	Ingestion, Storage, Analytics, Policies

Table 2.2: Use Case 1 non-functional requirements

2.2 Innovation: Protection mechanisms and tools

When we considered the requirements presented in the previous section, and developed the architecture for the platform, we integrated and enhanced a number of open source tools as well as developed our own tools and techniques. These represent our innovation from MOSAICrOWN which we continue to develop for exploitation within EISI. In order to fully understand the tools developed, we encourage the consultation of Chapter 2 of Deliverable D2.2 “Report on Requirements, Research Alignment and Deployment Plan” and Chapter 1 of Deliverable D2.3 “Final Report on Research Alignment”.

For context within this document, we provide a brief overview of the mechanisms and tools developed during the lifetime of MOSAICrOWN. EISI was primarily focused on the ingestion phase of the data lifecycle and, as such, had particular interest in the semantic models and vocabularies associated with the use case data, i.e., ICV and EV charging station data. Therefore, alongside the protection mechanisms, we needed to investigate and develop a platform for this ingestion and semantification such that the data is usable within a semantic toolchain. Indeed, the platform developed to satisfy UC1 is primarily constructed of open source tools with enhancements as deemed necessary to satisfy the UC1 requirements.

To understand the innovations from UC1, we can align the components of the platform with the dimensions of the requirements. To support the Data Ingestion (DI) and Data Governance (DG) requirements we developed an Android Auto application to facilitate transfer of ICV data and allow policy selection. Further support for DI requirements it provided by the Apache NiFi data flow which facilitated a “no code/low code” approach. Worth mentioning is the NiFi processor developed by EISI to facilitate JSON-LD to RDF triples conversions. For Access Control Management (ACM) and Data Processing (DP) requirements we developed a customer facing user interface which controls access and also analyzes data accessed from the data market. Data Management (DM) requirements are satisfied through a combination of components.

As for non-functional requirements, Data Economy (DE) requirements are primarily satisfied through the WebUI component, while Performance (P) and Code Quality (CQ) requirements were considered throughout the project.

2.3 Evaluation

In the following subsections we discuss how the requirements map to our tools (Section 2.3.1). We also provide an assessment of the tools and how they interact with UC1 (Section 2.3.2).

2.3.1 Evaluation of requirements

We consider the requirements as described in Deliverable D2.1 "Requirements from the Use Cases" and how our platform and the components therein satisfy those requirements. In Table 2.3 we show which component(s) of the platform satisfy the functional requirements. In Table 2.4 we show which component(s) of the platform satisfy the non-functional requirements. It should be noted that some of the requirements are only satisfied by the overall platform rather than individual components.

Requirement Reference	Description	Covered by Component
REQ-UC1-DI1	Close to source deployment	Android Auto application
REQ-UC1-DI2	Real-time stream handling	Apache Nifi
REQ-UC1-DI3	Batch handling	Apache NiFi
REQ-UC1-DI4	Support for different data types, structured and unstructured data	Apache NiFi
REQ-UC1-DI5	Data wrapping and data sanitization	Data Market Filter, FreyaFS
REQ-UC1-DI6	Compression and Encryption	TLS for data in flight, FreyaFS, HDFS
REQ-UC1-DI7	Secure identifier preservation	Policy Engine, Data Market Filter
REQ-UC1-DI8	Assessment of data completeness	Data Market Filter
REQ-UC1-DI9	Protection from linkage attacks	Data Market Filter
REQ-UC1-DI10	Support ingestion from multiple concurrent sources	Apache NiFi
REQ-UC1-DG1	Language and definitions for data governance	Policy Language
REQ-UC1-DG2	Support data governance models per data set per data provider	ICV Schema, Policy Language
REQ-UC1-DG3	Wrapping and sanitization parameters configurable by data owner	ICV Schema, Android Auto application, Policy Language
REQ-UC1-AC1	Access control and authorization	WebUI, Policy Engine, Policy Language
REQ-UC1-AC2	Grant and revoke access	WebUI, Policy Engine, Policy Language
REQ-UC1-AC3	Centralized key management infrastructure	PyKMIP, FreyaFS
REQ-UC1-AC4	Support limitation to data sharing or analytics	Policy Engine, Policy Language
REQ-UC1-AC5	Allow data sharing between providers and consumers	WebUI, Policy Engine, Policy Language

REQ-UC1-AC6	Allow data sharing between multiple parties	WebUI, Policy Engine, Policy Language
REQ-UC1-AC7	Policies configurable by data provider	Android Auto, Policy Language
REQ-UC1-DM1	Tracking data movement and access	Apache NiFi
REQ-UC1-DM2	Accessibility of data	Apache Jena, Apache Hadoop
REQ-UC1-DM3	Integrity of original data	Apache Hadoop, Apache NiFi
REQ-UC1-DM4	Support for deletion guarantees	FreyaFS
REQ-UC1-DM5	Protection at rest and in transfer	FreyaFS, TLS
REQ-UC1-DP1	Support storing data analytics results	WebUI, Apache NiFi
REQ-UC1-DP2	Anonymization of data analytics results	Apache NiFi, Data Market Filter
REQ-UC1-DP3	Merge data analytics results with shared data	WebUI, Apache NiFi

Table 2.3: Use Case 1 functional requirements and corresponding tools

The license model (REQ-UC1-DE1) is primarily satisfied by the platform in the form of customer interaction via the WebUI however all the individual components are required to enforce the licensing model.

Requirement Reference	**Description**	**Covered by Component**
REQ-UC1-DE1	License model	Platform
REQ-UC1-DE2	Distinction between data sharing and analytics	WebUI
REQ-UC1-P1	Limiting latency caused by wrapping and sanitization	Data Market Filter, FreyaFS
REQ-UC1-P2	Ensure utility under sanitization	WebUI
REQ-UC1-CQ1	Consider recommended coding practices	Platform
REQ-UC1-CQ2	Code covering for testing	Platform

Table 2.4: Use Case 1 non-functional requirements and corresponding tools

2.3.2 Assessment

UC1 aims to directly evaluate the data ingestion to the platform, taking into account the application of data governance policy, data wrapping techniques, and data sanitization techniques. In a more concrete sense, from the perspective of the actors in UC1, we set out to enable the connected vehicle fleet manager and the EV charging infrastructure provider the ability to exchange data such

that they can both derive benefit from the transaction; the vehicle fleet manager is able to monetize their data in a safe and secure way, and the EV charging provider gains valuable insights into how their infrastructure is being utilized. To achieve this goal, we developed tools and techniques to ingest this data, facilitate annotation of the data with privacy policy, treat and prepare this data for deployment in a data market, and finally control access to that data and analytics such that the privacy of the data owners is protected while still providing useful information. In this subsection, we provide a summary of this assessment.

The application of data governance required the most significant investment in thought, tools, and techniques. Data governance must permeate every facet of the platform. To achieve this, we centered our design around a flexible and extensible messaging broker, Apache NiFi, which allowed us to semantify the data as well as routing the data to various components of the platform for further treating. Also crucial in this design was the knowledge graph, Apache Jena with Apache Jena Fuseki, which allowed us to represent our metadata and policy in a linked fashion and also facilitated the separation of the data from the metadata. Also, tightly coupled together, are the front end web based user interface and the policy engine. The application of data governance policy was facilitated by leveraging the tools developed by the academic partners within WP3.

Data wrapping was another aspect of UC1 where we leveraged tools developed by academic partners within WP4. We successfully integrated the All-Or-Nothing Transform based encrypted filesystem into our platform. Indeed, we successfully containerized the filesystem and integrated it with other EISI tools such as Elastic Cloud Storage (ECS). Also, the data market filter component was required for both data wrapping and data sanitization, allowing us to treat the data prior to it's arrival at the data market.

For data sanitization we relied primarily on policy application to redact information as deemed appropriate by the data owner. The limitation of time hindered a more complete development of the data sanitization features.

All the components combined provided an effective platform for enabling the UC1 actors in fulfilling analytical needs as well as providing data owners, that is, the vehicle fleet manager or the driver of the vehicle (depending on the context of the data collection), with mechanisms to protect their data as well as with opportunities to trade personal data for monetary gain. Our work enables an economic model for sustaining the MOSAICrOWN data market.

3. Use Case 2 (MC)

Use Case 2 (UC2) considers financial institutions in the context of confidentially and confidently sharing data. Financial institutions manage a wide diversity of data which includes different levels of personally identifiable information (PII). Financial institutions need to understand how to define, store, wrap and analyze this PII. Without the ability to securely store, access and utilize this data, financial institutions lose a key part of their revenue and innovation streams, with a critical impact in their credibility and the enforcement of the privacy regulations.

As the regulatory and compliance landscapes change and data is continuously shared amongst multiple parties, the usage of personal data has increasingly narrowed. Companies needs to find appropriate ways to respect the regulations adapting to local regulations and while being innovative. Indeed, organizations must develop effective data strategies and utilize privacy enhancing techniques to meet regulation requirements, consumer expectations and to continue to innovate.

In this context Mastercard (MC) is responsible for UC2 which is showing the capacity of the POC which covers different dimensions of MOSAICrOWN. UC2 covers the ingestion, storage and analytics of the data, focusing on the sanitization and wrapping techniques in alignment with the privacy policies within MOSAICrOWN.

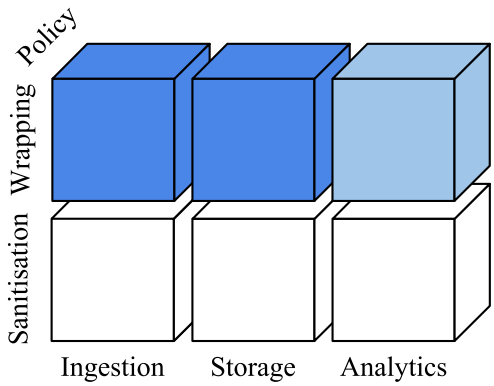


Figure 3.1: Dimensions covered by Use Case 2

The MC proof-of-concept platform is designed to answer different user needs through two different user journeys. One user journey to generate anonymized data applying wrapping techniques and another one to enable the visualization of analytics using the anonymized data.

In the remainder of this chapter, we first detail the requirements considered for this use case in Section 3.1. In Section 3.2, we detail the innovations produced during MOSAICrOWN to realize the use case. In Section 3.3, we evaluate how the developed techniques and tools address the requirements and assess their fit for practical applications.

3.1 Requirements

In this section, we overview the requirements to realize UC2 as defined in Deliverable D2.1 "Requirements from the Use Cases". Table 3.1 and Table 3.2 list the functional and non-functional requirements, respectively, with a short description and the dimension covered by the requirement. The functional requirements can be grouped in five aspects:

- **Data Ingestion (DI)**
- **Access Control Management (AC)**
- **Data Management (DM)**
- **Data Wrapping (W)**
- **Sanitization (S)**

Requirement Reference	Description	Dimension
REQ-UC2-S1	Data checks	Sanitization
REQ-UC2-S2	Unique Outlier	Sanitization
REQ-UC2-W1	Data Risk Assessment	Data Wrapping
REQ-UC2-W2	Data Wrapping Approach	Data Wrapping
REQ-UC2-W3	Data Catalogue	Data Wrapping
REQ-UC2-W4	Wrapping Techniques	Data Wrapping
REQ-UC2-W5	Pseudonymization	Data Wrapping
REQ-UC2-W6	Re-anonymization for UIDs	Data Wrapping
REQ-UC2-W7	Data linkage protection	Data Wrapping
REQ-UC2-W8	Encryption	Data Wrapping
REQ-UC2-W9	Encryption algorithms	Data Wrapping
REQ-UC2-W10	Encryption Keys	Data Wrapping
REQ-UC2-W11	Hashing	Data Wrapping
REQ-UC2-W12	Salting	Data Wrapping
REQ-UC2-W13	Rainbow Tables	Data Wrapping
REQ-UC2-W14	Tokenization	Data Wrapping
REQ-UC2-W15	Token Vault	Data Wrapping
REQ-UC2-W16	Horizontal Data Fragmentation	Data Wrapping
REQ-UC2-W17	Vertical Data Fragmentation	Data Wrapping
REQ-UC2-W18	Hybrid Data Fragmentation	Data Wrapping
REQ-UC2-AC1	Access control levels	Storage
REQ-UC2-AC2	User level access control	Storage
REQ-UC2-AC3	Business/Org level access control	Storage
REQ-UC2-AC4	API Access only	Storage
REQ-UC2-AC5	Permissible Purpose	Storage
REQ-UC2-AC6	Data export - aggregates only	Storage
REQ-UC2-DI1	Batch handling	Ingestion

REQ-UC2-DI2	Data types and formats	Ingestion
REQ-UC2-DI3	Data feeds via APIs	Ingestion
REQ-UC2-DI4	Data perturbation	Ingestion
REQ-UC2-DM1	Data controller rights	Storage
REQ-UC2-DM2	Controller to Controller Contract	Storage
REQ-UC2-DM3	Data combination and merging	Storage, Analytics
REQ-UC2-DM4	Limit data combinations	Storage, Analytics

Table 3.1: Use Case 2 functional requirements

The non-functional requirements cover two aspects:

- **Data Controller (DC)**
- **Data Analytics (DA)**

Requirement Reference	Description	Dimension
REQ-UC2-DC1	Data controller	Storage, Analytics
REQ-UC2-DC2	External unique identifiers	Storage, Analytics
REQ-UC2-DA1	Data science libraries	Storage, Analytics
REQ-UC2-DA2	Service SLAs	Storage, Analytics

Table 3.2: Use Case 2 non-functional requirements

3.2 Innovation: Protection mechanisms and tools

UC2 is designed to allow the platform user, as part of a broader organization, to anonymize data and to generate analytics leveraging the anonymized data. The overall objective is to support different policies, which call upon a set customized of wrapping techniques adapted to data type and local regulation. Indeed, the policy language enables the expression of privacy regulations. The user also has the possibility to select different wrapping techniques to be applied to the dataset. The functional requirements of UC2 are provided in Table 3.1.

The technology for UC2 is based on a Web Application where each user can access the platform, upload their dataset, and generate analytics after data anonymization. Indeed, UC2 is a cloud-based platform. The presentation tier communicates with the application tier and allows users to authenticate (Identity layer). The access to the application is done through the web browser, and the access is managed by an Identity layer cloud component allowing users to access the platform with the related grants, according with their user rights.

UC2 has the following objectives:

- User uploads a dataset to be anonymized in order to remove any PII.

- User selects a policy, corresponding to a set of wrapping techniques defined according to the local regulation requirements.
- The platform runs a semantic analysis of the data provided, recognizing data type, semantic and data distribution of the dataset uploaded by the user.
- User applies wrapping techniques pre-defined by the policy selected at the beginning of the user journey.
- User has the choice to customized the wrapping techniques pre-defined by the policy by selecting different wrapping techniques proposed in the drop-down menu.
- User uploads a fully anonymized file as a final output.
- User visualizes analytics generated from fully anonymized data.

UC2 is industry and privacy regulation independent. It is industry independent since the solution can take care of any type of data (account, order, ...) coming from any industry (Telecommunications, FSI, media, utility, healthcare,...) generating for each of them an anonymized dataset and the related insights (based on a market analytics template). It is privacy regulation independent since it can manage any type of regulation (GDPR for Europe, POPI for South Africa, LGPD for Brasil, ...) modeled into the application using the policy language provided by the MOSAICrOWN framework.

UC2 is fully integrated into the MOSAICrOWN project thanks to the versatility of the proof-of-concept platform. Indeed, the integration of a policy matching the MOSAICrOWN format (an extension of ODRL) produces a JSON file to complete the set of pre-defined data wrapping techniques to apply to the dataset uploaded by the user.

Academic partners of MOSAICrOWN have developed within WP3 a policy engine that has been integrated via a REST API with a predefined authorization method (an admin user capable to send a request via API). Inside the API, there is the possibility to load a JSON file containing the new policy specifications that we want to add in MOSAICrOWN platform. This JSON file must respect a particular format in order to have a fully functional policy specification in terms of data semantic types and possible data wrapping techniques to be applied.

Inside the platform, there is also a section called “Analytics” which is able to show some insights and KPIs calculated directly from anonymized datasets that users upload to the platform. At the moment this section has been customized for a particular vertical (Financial) to show its capabilities and potential.

3.3 Evaluation

3.3.1 Evaluation of requirements

The tables of requirements (Table 3.3 and Table 3.4) consider the requirements as described in Deliverable D2.1 “Requirements from the Use Cases” and their implementation status in our platform. In these tables, we use the terms *platform*, *production*, and *planned*, with the following meaning:

- Platform: the POC tool built within MOSAICrOWN supports the requirement.

- Production: the requirement is included in the design and will be available in the version of the product released to customers.
- Planned: the capability will be introduced in future releases.

Requirement Reference	Description	Covered by Component
REQ-UC2-DI1	Batch handling	Platform
REQ-UC2-DI2	Data types and formats	Platform
REQ-UC2-DI3	Data feeds via APIs	Platform
REQ-UC2-DI4	Data perturbation	Platform
REQ-UC2-AC1	API Access only	Production
REQ-UC2-AC2	User level access control management system	Production
REQ-UC2-AC3	Organization level access control management system	Production
REQ-UC2-AC4	API Access to all data	Planned
REQ-UC2-AC5	API Access Permission	Planned
REQ-UC2-AC6	Pseudonymization Data	Platform
REQ-UC2-DM1	Data handling inside platform	Platform
REQ-UC2-DM2	Controller to Controller Contract	Planned
REQ-UC2-DM3	Data combination and merging	Planned
REQ-UC2-DM4	Data upload limitation	Platform
REQ-UC2-W1	Data Semantic identification	Platform
REQ-UC2-W2	Data wrapping techniques association	Platform
REQ-UC2-W3	Data Wrapping selection	Platform
REQ-UC2-W4	Combination of data Wrapping techniques	Platform
REQ-UC2-W5	Data platform management	Platform
REQ-UC2-W6	Data upload	Platform
REQ-UC2-W7	Encryption algorithms	Platform
REQ-UC2-W8	Use of multiple encryption algorithms in data pipelines	Planned
REQ-UC2-W9	Encryption keys	Platform
REQ-UC2-W10	Hashing algorithms	Platform
REQ-UC2-W11	Application of a salt to the hashing algorithm chosen	Planned
REQ-UC2-W12	Rainbow tables	Production
REQ-UC2-W13	Tokenization	Production
REQ-UC2-W14	Storage of the generated tokens and their corresponding data in an encrypted token vault	Planned
REQ-UC2-W15	Support horizontal data fragmentation on demand	Planned
REQ-UC2-W16	Support vertical data fragmentation when storing the data on demand	Planned
REQ-UC2-W17	Support hybrid data fragmentation on demand	Planned
REQ-UC2-S1	Data control during the upload and during data wrapping techniques application	Platform
REQ-UC2-S2	Remove unique outliers in dataset as data is loaded	Planned

Table 3.3: Use Case 2 functional requirements and implementation status

Requirement Reference	Description	Covered by Component
REQ-UC2-DC1	Data Repository	Platform
REQ-UC2-DC2	External unique identifiers	Planned
REQ-UC2-DA1	KPIs and insights show on Data Analytics section	Platform
REQ-UC2-DA2	Service SLAs	Production

Table 3.4: Use Case 2 non-functional requirements and implementation status

3.3.2 Assessment

The proof-of-concept consists of three applications, one for each functionality of the platform. The core of the proof-of-concept is based on Data Semantic / Data Wrapping (DSDW) Engine and its functionalities (for a detail description, refer to Deliverable D2.4 "Use Case Prototypes").

Once the dataset is received from the user, the same dataset is saved in the bucket named with a unique alphanumeric identifier. A new job is created in the job table associated with that temporary file. The DSDW Engine takes over the job and performs a pre-processing phase on the received dataset based on tokenization/embedding extraction processes (required for the next phases). After that phase the DSDW Engine computes, for each column of the dataset, the probability that this column belongs to a particular semantic type according to its content. The semantic type with the highest probability is selected. Once the semantic type of each column has been identified, the job status is updated to allow the interface to show the result of this analysis.

Once the user has selected for each column the data wrapping technique to be applied, the DSDW Engine applies these techniques on the original dataset creating a new temporary file (containing the data in anonymized form). This temporary file is then made available to the user who will have the possibility to download it.

Data Analytics

The Analytics section displays a dashboard that is able to collect all correctly loaded data showing certain KPIs and insights inherent to financial domain such as average spend per ticket, number of transactions and total spend.

During the upload phase the user has the possibility to select if the uploaded dataset is eligible to be used inside Data Analytics section. The platform is capable to understand if the provided dataset respects the requested format (vertical FSI). If the dataset respects that format, the platform uploads the anonymized version of that dataset into the data analytics section. The user can also download this format as a template in order to understand which dataset format is accepted.

Inside the Data Analytics section the user can see some KPIs (e.g., analytics on performance of product, channel, card type and category of purchases with a monthly, quarterly or yearly view) and insights related to Financial vertical in an aggregated and pseudo-anonymized way.

Process diagram

The process of data analysis and wrapping, starts from an unstructured dataset provided by the user. The process is fully asynchronous (mandatory considering the amount of data that could be involved in this data transformation) so the user can run it and then retrieve the related job monitoring the execution or moving to the next step. The same approach is used for the application of the selected data wrapping techniques considering also the possibility to upload the anonymized dataset on analytics database if the dataset format is correct.

Sequence process

The platform has several layers related to the anonymization process. Regarding the Analytics section we define an API layer that receives Section, StartDate and EndDate as parameters and provides the relative data to populate the requested section in terms of dimensions, KPIs and graphs.

4. Use Case 3 (SAP SE)

This chapter reports on the final evaluation of Use Case 3 (UC3) with regards to how the tools developed during MOSAICrOWN align with the requirements defined in Deliverable D2.1 "Requirements from the Use Cases".

Next, we briefly recall the use case description. UC3 considers privacy-preserving analytics in a business-to-business context. In more detail, businesses with common interests cannot easily share business-sensitive information (operational data) or customer-generated data (experience data) but their combined data contains valuable insights to foster collaboration.

For example, producers and retailers aiming to better allocate marketing budgets and improve their distribution process. Figure 4.1 overviews the supported ingestion considered in UC3. Data can be shared directly in plaintext and combined with other data before being sanitized, or data can be shared after sanitization, or data can be wrapped, i.e., encrypted suitable for secure computation, to perform a distributed analysis. When the data is sanitized during ingestion, we call this the *local model*. When the data is sanitized before analytics, i.e., stored in plaintext, we call this the *central model*. The *hybrid model* uses additional cryptographic techniques in the form of secure computation to directly perform the analytics step in a distributed manner where the data is protected via a form of encryption.

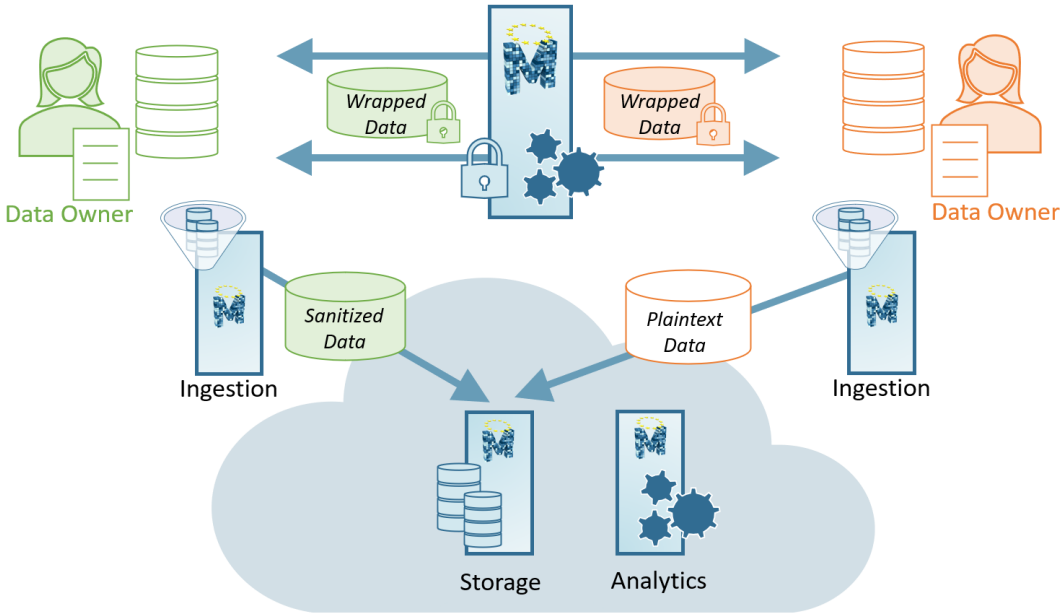


Figure 4.1: Flexible data sharing: plaintext data, sanitized data, or wrapped data suitable for secure computation

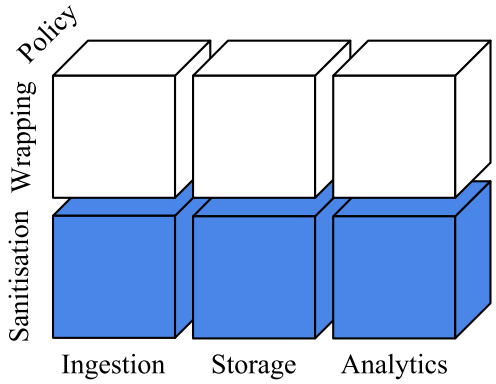


Figure 4.2: Dimensions covered by Use Case 3

UC3 is concerned with the entire data lifecycle for sanitization, covering all dimensions; namely, ingestion, storage, and analytics as visualized in Figure 4.2.

In the remainder of this chapter, we first detail the requirements considered for Use Case 3 in Section 4.1. In Section 4.2, we detail the innovations produced during MOSAICrOWN to realize the use case. In Section 4.3, we evaluate how the developed techniques and tools address the requirements and assess their fit for practical applications.

4.1 Requirements

In this section, we overview the requirements to realize UC3 as defined in Deliverable D2.1 "Requirements from the Use Cases". Table 4.1 lists the functional requirements with a short description and the dimension covered by the requirement. The functional requirements can be grouped in three aspects as summarized next.

Access control (AC) to ensure that only authorized users or groups can access the stored data. Data owners can define and enforce access rights via different means, e.g., encryption. To share data, the parties with the appropriate rights can grant access to the data of other parties.

Local function sanitization (SL) at time of ingestion in the cloud platform with customizable parametrization supporting the storage of sanitized data and the sharing of sanitized data.

Central function sanitization (SC) during the analytical function evaluation by the cloud platform with customizable parametrization. This aspect also considers secure computation between the data owners to compute sanitized analytics in a distributed fashion.

Requirement Reference	Description	Dimension
REQ-UC3-AC1	Access control per data owner	Storage
REQ-UC3-AC2	Access Control for data owner groups	Storage
REQ-UC3-AC3	Key per data owner	Storage
REQ-UC3-AC4	Keys for data owner groups	Storage
REQ-UC3-SL1	Local anonymization parameters chosen by the data owner	Ingestion, Sanitization, Policies

REQ-UC3-SL2	Storing the result of the sanitization service in the Cloud	Ingestion, Sanitization
REQ-UC3-SL3	Sharing sanitized results with other data owners	Storage, Policies
REQ-UC3-SC1	Central anonymization parameters chosen by the data owner	Analytics, Sanitization, Policies
REQ-UC3-SC2	Collecting inputs from multiple data owners for aggregation	Analytics, Sanitization
REQ-UC3-SC3	Collecting inputs from multiple data owners via secure computation	Ingestion, Analytics, Sanitization
REQ-UC3-SC4	Anonymization to protect identity of data subjects and hinder re-identification	Ingestion, Analytics, Sanitization

Table 4.1: Use Case 3 functional requirements

Table 4.2 lists the non-functional requirements with a short description and the dimension covered by the requirement. The non-functional requirements given in the table cover three aspects as detailed in the following.

Performance (P) of the privacy-preserving analysis functionalities, i.e., with the application of sanitization techniques, should be acceptable for practical evaluations.

Extendability (EX) should be supported to allow flexible and simple addition of further sanitization techniques to reflect the current state-of-the-art and provide sound technical means.

Interpretability (IN) for anonymization guarantees for a given parametrization should be provided by quantifying the risk associated with re-identification attacks.

Requirement Reference	**Description**	**Dimension**
REQ-UC3-P1	Scalability of anonymization functionalities	Ingestion, Analytics, Sanitization
REQ-UC3-P2	Utility maximizing anonymization functionalities	Ingestion, Analytics, Sanitization
REQ-UC3-EX1	Extendability pattern for anonymization functions	Ingestion, Analytics, Sanitization
REQ-UC3-EX2	Cloud platform provider should provide sound technical means for anonymization	Ingestion, Analytics, Sanitization

REQ-UC3-IN1	Anonymization services should be accompanied by simulated adversary	Ingestion, Analytics, Sanitization

Table 4.2: Use Case 3 non-functional requirements

4.2 Innovation: Protection mechanisms and tools

To address the requirements, we developed different tools and techniques, which represent the innovation achieved during MOSAICrOWN. For a technical overview and background information regarding the sanitization mechanisms satisfying differential privacy as well as the secure computation techniques used by our tools, we refer to Section 3.2 in Deliverable D2.4 “Use Case Prototypes”.

In more detail, during MOSAICrOWN we investigated privacy metrics, especially the interpretation and quantification of privacy guarantees, via an attack-based framework. Namely, membership inference attacks, which aim to inference the inclusion of an individual in training data used in the context of machine learning applications. Our tool MIA provides such attack simulations and evaluations of the privacy guarantee (covering requirement REQ-UC3-IN1). The technological architecture overview for MIA is detailed in Deliverable D5.1 “First version of data sanitisation tools” and the theoretical background is presented in Deliverable D5.2 “First report on privacy metrics and data sanitisation” and extended in Deliverable D5.3 “Final report on privacy metrics, risks, and utility”.

Furthermore, our tool DPtool provides an intuitive graphical user interface accustomed to business users (SAPUI5) supporting multiple differentially private mechanisms and customizable parametrization. Also, DPtool exposes a REST API for programmatic access. The architecture and considered technology stack for DPtool is detailed in Deliverable D2.3 “Final Report on Research Alignment” and D2.4 “Use Case Prototypes”.

Our tool DPsc enables multiple, distributed parties to compute a differentially private statistic over their joint data without revealing their data to each other. More precisely, DPsc uses secure computation to compute differentially private rank-based statistics (i.e., percentiles), such as the median. The theoretical background and formal description of DPsc is given in Deliverable D5.4 “Final versions of tools for data sanitisation and WP5 computation” and extended in Deliverable D5.5 “Report on data sanitisation and computation”. The technological architecture and usage of DPsc is described in Deliverable D2.4 “Use Case Prototypes”.

4.3 Evaluation

In the following, we evaluate the requirements and perform an assessment with regards to our presented tools. In more detail, we describe how the requirements map to our tools in Section 4.3.1 and provide an assessment of the tools in an operational situation in Section 4.3.2.

4.3.1 Evaluation of requirements

Our tools fully meet the functional requirements defined in Deliverable D2.1 “Requirements from the Use Cases”. DPtool addresses local sanitization requirements REQ-UC3-SL1, REQ-UC3-SL2 as well as central sanitization requirements REQ-UC3-SC1, REQ-UC3-SC4. MIA also considers central parametrization, i.e., REQ-UC3-SC1. DPsc addresses the sanitization requirements not already met by DPtool, namely, REQ-UC3-SC2 and REC-UC3-SC3.

Table 4.3 lists the tools and which function requirements they address. Note that access control requirements (REQ-UC3-AC1, REQ-UC3-AC2, REQ-UC3-AC3) are not listed as they are met by existing solutions (e.g., SAP HANA).

Requirement Reference	Description	Covered by Tool
REQ-UC3-SL1	Local anonymization parameters chosen by the data owner	DPtool
REQ-UC3-SL2	Storing the result of the sanitization service in the Cloud	DPtool
REQ-UC3-SL3	Sharing sanitized results with other data owners	DPsc
REQ-UC3-SC1	Central anonymization parameters chosen by the data owner	DPtool, MIA
REQ-UC3-SC2	Collecting inputs from multiple data owners for aggregation	DPsc
REQ-UC3-SC3	Collecting inputs from multiple data owners via secure computation	DPsc
REQ-UC3-SC4	Anonymization to protect identity of data subjects and hinder re-identification	DPtool

Table 4.3: Use Case 3 functional requirements and corresponding tools

Also, our tools fully address the non-functional requirements. All our tools satisfy differential privacy, a rigorous and formal anonymization notion, providing sound technical means, i.e., our tools address REQ-UC3-EX2. DPtool covers performance requirements REQ-UC3-P1 and REQ-UC3-P2 and supports extensions of the provided sanitization mechanism, i.e., REQ-UC3-EX1. DPsc covers the same requirements as DPtool, however, as DPsc was designed for efficiency with secure computation it is not as extendable as DPtool. Finally, MIA covers REQ-UC3-IN1. Table 4.4 lists the non-functional requirements and the corresponding tools satisfying the requirement.

Requirement Reference	Description	Covered by Tool
REQ-UC3-P1	Scalability of anonymization functionalities	DPtool, DPsc
REQ-UC3-P2	Utility maximizing anonymization functionalities	DPtool, DPsc
REQ-UC3-EX1	Extendability pattern for anonymization functions	DPtool
REQ-UC3-EX2	Cloud platform provider should provide sound technical means for anonymization	DPtool, DPsc, MIA

REQ-UC3-IN1	Anonymization services should be accompanied by simulated adversary	MIA

Table 4.4: Use Case 3 non-functional requirements and corresponding tools

4.3.2 Assessment

Our techniques and tools provide state-of-the-art protection via sanitization as well as secure computation with practical performance suitable for the privacy-preserving analytics use case.

In more detail, all our tools satisfy differential privacy, a strong privacy notion, which is currently applied in a wide range of industry scenarios [App16, App17, DKY17, Rog20, RSP+21] and by the US census bureau [Abo18].

For membership inference attacks, we quantified privacy risks thoroughly in Deliverable D5.1 "First version of data sanitisation tools" for different machine learning applications (i.e., feed-forward neural networks and generative models). Our tool MIA is built on top of TensorFlow [ABC+16] which is commonly used by data scientists for deep neural networks and is widely deployed in industry settings.[1] As such, the performance of our tool mainly relies on the underlying implementation for machine learning models as provided by TensorFlow, which is sufficient for industry use cases and applications as evident from its wide-ranging applications in this space.

For general sanitization, the REST API of our tool DPtool implements multiple differential privacy mechanisms in Java. Mechanisms based on additive noise (e.g., Laplace mechanism) provide more than suitable performance (requiring milliseconds for thousands of records).

For secure computation of differentially private statistics, our tool DPsc is optimized for differentially private rank-based statistics (i.e., percentiles) and built with the ABY framework [DSZ15]. The ABY framework provides efficient secure computation for different paradigms (garbled circuits and secret sharing, see Deliverable D2.4 "Use Case Prototypes" for details) as well as conversions between them and is actively maintained[2], i.e., receiving further improvements and optimizations. While general solutions for secure computation of differential private statistics exist in theory (i.e., by straightforward implementation with general-purpose secure computation), such solutions require large computation overhead and are not practical. Our tool DPsc, however, achieves running times of few seconds even in a wide-area network (where the main bottleneck is the network delay), as detailed in Deliverable D5.4 "Final versions of tools for data sanitisation and WP5 computation". The running time of DPsc is around an order of magnitude faster than comparable tailor-made solutions and provides higher accuracy (see evaluations in Deliverable D5.4 "Final versions of tools for data sanitisation and WP5 computation" and D5.5 "Report on data sanitisation and computation").

Altogether, our tools provide suitable performance for practical applications and offer strong protection via state-of-the-art sanitization as well as cryptographic techniques. Note that performance improvements in the underlying libraries (i.e., machine learning library TensorFlow and secure computation framework ABY) directly improve our applications as well, helping us to leverage advancement in state-of-the-art with respect to these libraries.

[1] https://www.tensorflow.org/about/case-studies
[2] https://github.com/encryptogroup/ABY

5. Conclusions

This deliverable provided an evaluation of the prototypes that were developed in each MOSAICrOWN use case by the industrial partners, and were presented in Deliverable D2.4 “Use Case Prototypes”. This evaluation considers each one of the functional and non-functional requirements that were identified and presented in Deliverable D2.1 “Requirements from the Use Cases”, and aligns them with the innovations developed in each use case.

Chapter 2 presented the evaluation for Use Case 1, focusing on Intelligent Connected Vehicles. The platform, developed by EISI, relies on a number of open-source components. In order to satisfy all the requirements identified in Deliverable D2.1, some of them had to be adapted or extended. EISI was also able to successfully reuse some components developed by other partners of the project.

Chapter 3 presented the evaluation for Use Case 2. This use case focuses on transaction-level data analytics for financial institutions, but the web-based platform developed by MC could easily be extended to support other domains. Similarly, while the platform currently supports two privacy regulations (Europe’s GDPR and Brazil’s LGPD), it could easily be adapted to other regulations, thanks to the expressiveness of the policy language. The chapter reported the implementation status of the requirements identified in Deliverable D2.1, and discussed further developments.

Chapter 4 presented the evaluation for Use Case 3, which focuses on sanitization at every stage of the data lifecycle. A suite of tools have been developed by SAP SE that apply differential privacy, a strong and formally defined privacy notion, largely adopted in the scientific literature and by the industry. The chapter discussed how each of the functional and non-functional requirements identified in Deliverable D2.1 are addressed by at least one of the tools in this suite.

Bibliography

[ABC+16] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. Tensorflow: A system for large-scale machine learning. In *Proc. of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, Savannah, GA, USA, November 2016.

[Abo18] J.M. Abowd. The U.S. Census Bureau Adopts Differential Privacy. In *Proc. of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, London, U.K., August 2018.

[App16] Apple. WWDC 2016: Engineering Privacy for Your Users, 2016. `https://developer.apple.com/videos/play/wwdc2016/709/`.

[App17] Apple. Apple's differential privacy team: Learning with privacy at scale, 2017. `https://machinelearning.apple.com/2017/12/06/learning-with-privacy-at-scale.html`.

[Com20a] European Commission. Data Governance Act. Draft legislation COM(2020) 767 final, European Commission, November 2020.

[Com20b] European Commission. Digital Markets Act: Proposed Regulation on contestable and fair markets in the digital sector. Draft legislation 2020/842/COM, European Commission, December 2020.

[Com20c] European Commission. (Digital Services Act) Proposal for a Single Market For Digital Services and amending Directive 2000/31/EC. Draft legislation 2020/825/COM, European Commission, December 2020.

[Com20d] European Commission. A European strategy for data. COM 2020/66/COM final, European Commission, February 2020. Doc ID: 52020DC0066 Doc Sector: 5 Doc Title: COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS A European strategy for data Doc Type: DC Usr_lan: en.

[Com21] European Commission. Proposal for a Regulation on a European approach for Artificial Intelligence | Shaping Europe's digital future. Draft legislation EC-2020-206-COM, European Commission, April 2021.

[DKY17] B. Ding, J. Kulkarni, and S. Yekhanin. Collecting telemetry data privately. In *Proc. of the 31st Conference on Neural Information Processing Systems (NIPS)*, Long Beach, CA, USA, December 2017.

[DSZ15] D. Demmler, T. Schneider, and M. Zohner. ABY-A framework for efficient mixed-protocol secure two-party computation. In *Proc. of the Network and Distributed System Security (NDSS) Symposium*, San Diego, CA, USA, February 2015.

[Rog20] R. Rogers. A Differentially Private Data Analytics API at Scale. In *Proc. of the 2020 USENIX Conference on Privacy Engineering Practice and Respect (PEPR)*, October 2020.

[RSP+21] R. Rogers, S. Subramaniam, S. Peng, D. Durfee, S. Lee, S. K. Kancha, S. Sahay, and P. Ahammad. LinkedIn's Audience Engagements API: A Privacy Preserving Data Analytics System at Scale. *Journal of Privacy and Confidentiality*, 11(3), December 2021.

[Uni16] European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), May 2016. Legislative Body: EP, CONSIL.