# auto twin

## Data-driven method based on a process mining approach for Automated Digital Twin generation, operations, and maintenance in circular value chains

| Deliverable | |
|---|---|
| **D7.1 Data Management Plan** | |
| **Deliverable Lead** | UBI |
| **Deliverable Due Date** | 31/05/2023 |
| **Actual Submission Date** | 22/05/2023 |
| **Version** | 1.0 |

# Document Control Page

| Title | **D7.1 Data Management Plan** |
|---|---|
| **Editor** | UBITECH |
| **Contributors** | PMI |
| **Work Package** | WP7 – Project Management |
| **Description** | Initial version of the AUTO-TWIN Data Management Plan documenting:<br><br>• The details of the AUTO-TWIN data collection and data generation process and its relation to the objectives of the project<br><br>• The initial list of data elements of AUTO-TWIN project<br><br>• The applied FAIR (findable, accessible, interoperable and re-usable) principles on each data element<br><br>• The AUTO-TWIN strategy for the data management allocation of resources and data security<br><br>• The details of the ethical aspects of each data element extending the information documented in deliverable D8.1 |
| **Creation Date** | 07/02/2023 |
| **Type** | Report |
| **Language** | English |
| **Audience** | ☒ Public<br>☐ Confidential |
| **Review Status** | ☐ Draft<br>☒ WP leader accepted<br>☒ Coordinator accepted |
| **Action Requested** | ☐ to be revised by Partners<br>☐ for approval by the WP leader<br>☐ for approval by the Project Coordinator<br>☐ for acknowledgement by Partners |

# Document History

| Version | Date | Author(s) | Changes |
|---|---|---|---|

| 0.1 | 07/02/2023 | UBI | ToC |
|-----|------------|-----|-----|
| 0.2 | 15/03/2023 | UBI | Initial contributions on Sections 1, 2 and 3 |
| 0.3 | 30/03/2023 | UBI | Initial contributions on Sections 4, 5 and 6 |
| 0.4 | 13/04/2023 | UBI | Updated contributions in all sections |
| 0.5 | 04/05/2023 | UBI | Updated contributions in all sections |
| 0.7 | 08/05/2023 | UBI | Finalisation of sections 2, 3 and 4 |
| 0.8 | 10/05/2023 | UBI | Finalisation of sections 5 and 6 |
| 0.9 | 12/05/2023 | UBI | Version for internal review |
| 0.95 | 18/05/2023 | SYX | Internal review of D7.1 |
| 0.95 | 18/05/2023 | TTS | Internal review of D7.1 |
| 1.0 | 23/05/2023 | UBI | Final version |

## Table of Contents

# List of Tables

# Definitions and acronyms

| | |
|---|---|
| CSV | Comma Separated Values |
| DMP | Data Management Plan |
| DOI | Digital Object Identifier |
| EC | European Commission |
| EU | European Union |
| FAIR | Findable, Accessible, Interoperable and Re-usable |
| IDS | International Data Space |
| IPR | Intellectual Property Right |
| M | Month |
| N/A | Not Applicable |
| TBD | To be defined |
| WP | Work Package |

# Disclaimer

The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

## Executive Summary

The goal of Task 7.5 "Data protection and data management plan" is to define AUTO-TWIN Data Management Plan, which documents the core aspects of the data management policy of AUTO-TWIN as it will be applied by the consortium for all research data of the project.

This document presents the initial version of the Data Management Plan, documenting the list of data elements that will be collected, generated and processed during the project lifetime, providing a detailed documentation of each data element which covers:

- The scope of the data collection, the data generation process and its relation to the objectives of the project

- Their compliance with the FAIR principles in order to be findable, accessible, interoperable and re-usable to the highest possible degree following European Commission's guidelines

- The ethical aspects of each data element in accordance with the ethical rules and standards of Horizon Europe.

In addition to this, the Data Management Plan defines the details for the data management allocation of resources within the project as well as the designed and applied data security policies on all the project's research data.

The AUTO-TWIN Data Management Plan constitutes a living document in accordance with the guidelines of Horizon Europe hence it will be constantly enhanced and optimised during the project implementation in order to document the required updates from the advancements of the research activities of the project. Updates and improvements will be reported in Deliverable D7.2, expected at month 18.

# 1   Introduction

## 1.1   Description of the document

D7.1 is released in the scope of WP7 "Project Management" activities and documents the preliminary outcomes of Task 7.5 "Data protection and data management plan". The deliverable is prepared in accordance with the AUTO-TWIN Description of Action and reports the activities and results of the work performed as the first iteration within the context of Task 7.5. Task 7.5 started at month 1 and it will end in month 36.

The scope of the deliverable is to report the initial version of the AUTO-TWIN Data Management Plan (DMP). The scope of the AUTO-TWIN DMP is to provide the required analysis on the core elements of the data management policy which is applied in the course of the project by the AUTO-TWIN consortium with regard to the project research data.

The DMP was prepared in accordance to the "Guidelines on FAIR Data Management in Horizon 2020"[1] and it's built around the following main pillars based on these guidelines:

- the handling of research data during and after the end of the project

- what data will be collected, processed and/or generated

- which methodology and standards will be applied

- whether data will be shared/made open access

- how data will be curated and preserved (including after the end of the project).

which aim at making research data adhering to the FAIR (findable, accessible, interoperable and re-usable) principles.

DMP constitutes a living document, in accordance with the aforementioned guidelines, hence it will be optimised and updated with a finer level of granularity over the course of the project once significant changes arise as the project evolves such as the identification of new data and changes in consortium policies or composition.

The initial version of the AUTO-TWIN DMP is delivered in M6 as planned. While this first version does not contain detailed information of the AUTO-TWIN datasets which are collected, generated, and processed, curated, and persisted, it sets the foundations and principles of the data management policy of the project with regards to its research data during and after the end of the project. The forthcoming versions of the DMP will contain more details of the datasets and applied processes. To this end, the consortium will constantly monitor their evolution and introduce the required updates on the living document of the DMP.

## 1.2   WP and Tasks related with the deliverable

The deliverable is released in the scope of the WP7 "Project Management" and documents the work performed within the first period of the project till M6. The task is directly related to the rest of the Tasks of WP7, as it provides the means for the data protection processes of the project contributing to the overall project management activities while also taking input for the processes related to the quality and risk management as well as the societal and legal aspects.

In addition to this, the task is building on top of the outcomes of WP2 "Common Data Space for Circular Economies & Blockchain Platform Implementation ", WP3 "Models and Methods for Automated Generation & Update of Digital Twins for Circular Economies", WP4 "Enabling Green

---

[1] *https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf*

Gateways through Explainable Analytics & Decision Making" and WP6 "Communication and Impact" in which the various datasets are collected, generated, processed, curated and persisted in order to formulate and apply the appropriate data management policy of the project.

## 2   Data Summary

*It should be noted that the data management plan constitutes a living document following the guidelines of the EC, as mentioned before. Furthermore, at the moment of writing two of the project's use cases are still in prototyping phase thus the exact nature of the available data will be more clear after M12. Towards this end, the data management plan will be updated and optimised with additional and more detailed information following the advancements of the project during the complete project lifetime.*

AUTO-TWIN aims at providing a novel platform that will effectively address the technological shortcomings and economic liability of current system-engineering model towards the transition to a Circular Economy. The main innovations of the platform are: a) the introduction of an automated process-aware discovery towards the autonomous Digital Twin generation which supports the trustworthy business process in circular economies, b) the adoption of an (International Data Space) IDS-based common data space that promotes and facilitates the secure and seamless exchange of manufacturing /product/ business data within value-networks in a circular-economy ecosystem and finally c) the integration of novel hardware technologies into the digital thread with the aim of creating smart Green Gateways that will empower companies to perform data and digital twin enabled green decisions.

Towards this end, the particular section documents the information related to the data which are collected and processed during the project's duration elaborating on the purpose of the data collection and generation process and its relation to the objectives of the project.

Table 1 presents the complete list of data to be collected or generated as well as processed within the context of AUTO-TWIN project. In Section 2.1, each data element is documented in detail while in Section 3 the applied FAIR principles aspects of each data element are presented.

Table 1: AUTO-TWIN list of data elements

| # | Data Element Name | Data Element ID # | Data Element Owner | Origin WP |
|---|---|---|---|---|
| 1 | CSV files constructed from sensors | TUE_PM | TUE | WP2 (T2.1) |
| 2 | Twin Enriched GR3N Data | SYX_GR3N_TED | SYX | WP2 (T2.3) |
| 3 | Twin Enriched CROMA Data | SYX_CROMA_TED | SYX | WP2 (T2.3) |
| 4 | Twin Enriched Libattion Data | SYX_LIB_TED | SYX | WP2 (T2.3) |
| 5 | Process model data | PMI_PM | PMI | WP3 (T3.1) |
| 6 | Knowledge Graphs produced from sensors | TTS_DES | TTS | WP3 (T3.2) |
| 7 | System knowledge graph meta data | TUE_META | TUE | WP3 (T3.2, T3.3) |
| 8 | System knowledge graph data | TUE_GRAPH | TUE | WP3 (T3.2, T3.3) |

| 9 | Data from Archetypes Database | ERQ_ARC_DB | ERQ | WP4 (T4.4) |
|---|---|---|---|---|
| 10 | Data from Archetypes Macro Classes | ERQ_ARC_MCR_CLS | ERQ | WP4 (T4.4) |
| 11 | Data from Soft Skill Questionnaire | ERQ_SFT_SKL_QUE | ERQ | WP4 (T4.4) |
| 12 | Data from Archetypes by AI component | ERQ_ARC_AI | ERQ | WP4 (T4.4) |
| 13 | Data from Archetypes – Cluster - Tool | ERQ_ARC_CLS_TOOL | ERQ | WP4 (T4.4) |
| 14 | Project's Mailing List | COR_1_Newsletter | COR | WP6 (T6.1) |
| 15 | Social Media contacts & followers | COR_2_SocialMedia | COR | WP6 (T6.1) |
| 16 | Website cookies | COR_3_Website | COR | WP6 (T6.1) |
| 17 | Participants Lists | COR_4_EventsParticipants | COR | WP6 (T6.1) |

## 2.1 AUTO-TWIN data collection purpose and relation to objectives

Within the context of AUTO-TWIN project, on one hand data will be collected from the demonstrator partners while on the other hand new data will be generated. The purpose of the data collection is threefold and directly related to the activities of the project. In particular, AUTO-TWIN will collect data that will enable the discovery and replication of the business processes providing the basis for the process mining and system knowledge graph activities that will be performed towards the formulation of the simulation models for the automated Digital Twins generation for Circular Economies. Secondly, the collected data will enable the design and implementation of the aspired common data space which will facilitate the data exchange of industrial and manufacturing data in a secure and effortless manner. Finally, the collected data will drive the design and implementation of the process-mining, AI, explainability and profiling techniques which will enable the implementation of the smart green-gateways towards the formulation of valuable decision making in Circular Economies. In addition to the above, data will be collected and utilised as dissemination material of project in order to perform the required communication and dissemination activities.

To this end, AUTO-TWIN will collect, generate and process a variety of data elements which will be leveraged towards the realisation of the project objectives. In the following table, the details of each data element listed in Table 1 are presented, documenting in detail their origin, their types and formats, as well as their purpose and utilisation in the course of the project.

| Data Element ID | # TUE_PM |
|---|---|
| **Origin** | TUE will compare generated from original data in the graph, enriching and exporting the final data for Process Mining |
| **Type(s)** | <ul><li>Time-stamped data (e.g. state of equipment, from machines PLCs), Sensor location/Sensor ID (e.g. from IoT sensors on the equipment),</li><li>Estimated health-index (e.g. current lifetime of the equipment obtained from sensors)</li><li>Rank in the circular value chain (e.g. Current position)</li></ul> |

| Format(s) | • Comma Separated Values (.csv)<br>• XML<br>• OCEL<br>• Excel files (.xlsx) |
|---|---|
| Purpose | This type of data will help to identify the different type of processes criticalities that require an intervention to improve the process itself, perform conformance checking and identifying bottlenecks. |
| Utility | Different partners related to process mining tasks can use the data to apply different process mining algorithms and achieve crucial improvements in the course of the equipment within the circular value chain. |

| Data Element ID | # SYX_GR3N_TED |
|---|---|
| Origin | SYX will enrich and store GR3N pilot user data |
| Type(s) | • Time-stamped data<br>• Sensor location/Sensor ID<br>• Estimated health-index<br>• Rank in the circular value chain |
| Format(s) | • Comma Separated Values (.csv)<br>• JavaScript Object Notation (.json)<br>• Apache Cassandra database files (.db)<br>• MongoDB database files (.bson)<br>• Image files (.jpg, .png, .eps)<br>• Document files (.pdf, .docx, .pptx) |
| Purpose | The enriched data can help to provide more complete data to Automatic and Autonomous digital twins |
| Utility | The technical partners may use these data as a valid input to Automatic and Autonomous digital twins |

| Data Element ID | # SYX_CROMA_TED |
|---|---|
| Origin | SYX will enrich and store CROMA pilot user data |
| Type(s) | • Time-stamped data<br>• Sensor location/Sensor ID<br>• Estimated health-index<br>• Rank in the circular value chain |
| Format(s) | • Comma Separated Values (.csv)<br>• JavaScript Object Notation (.json)<br>• Apache Cassandra database files (.db)<br>• MongoDB database files (.bson)<br>• Image files (.jpg, .png, .eps)<br>• Document files (.pdf, .docx, .pptx) |
| Purpose | The enriched data can help to provide more complete data to Automatic and Autonomous digital twins |

| Utility | The technical partners may use these data as a valid input to Automatic and Autonomous digital twins |
|---|---|

| Data Element ID | # SYX_LIB_TED |
|---|---|
| Origin | SYX will enrich and store Libattion pilot user data |
| Type(s) | • Time-stamped data<br>• Sensor location/Sensor ID<br>• Estimated health-index<br>• Rank in the circular value chain |
| Format(s) | • Comma Separated Values (.csv)<br>• JavaScript Object Notation (.json)<br>• Apache Cassandra database files (.db)<br>• MongoDB database files (.bson)<br>• Image files (.jpg, .png, .eps)<br>• Document files (.pdf, .docx, .pptx) |
| Purpose | The enriched data can help to provide more complete data to Automatic and Autonomous digital twins |
| Utility | The technical partners may use these data as a valid input to Automatic and Autonomous digital twins |

| Data Element ID | PMI_PM |
|---|---|
| Origin | PMI will produce process model data for use cases or other demonstrators |
| Type(s) | • Input event logs<br>• Model configuration<br>• Model description<br>• Evaluation metrics |
| Format(s) | • Comma Separated Values (.csv)<br>• JavaScript Object Notation (.json)<br>• Neo4j database files (.db)<br>• Image files (.jpg, .png, .eps)<br>• Document files (.pdf, .docx, .pptx) |
| Purpose | The data can help to validate process discovery algorithms and develop the process discovery framework |
| Utility | The research community may use the data as a benchmark to compare different process discovery algorithms |

| Data Element ID | # TTS_DES |
|---|---|

| Origin | TTS generates data from "learned" simulation |
|---|---|
| Type(s) | Entities in the knowledge graphs (e.g. id for each equipment), Attributes (e.g. properties or characteristics of the entities), Relationships (e.g. connections between the entities in the knowledge graph) |
| Format(s) | • Comma Separated Values (.csv)<br>• XML<br>• JSON<br>• Excel files (.xlsx) |
| Purpose | This type of data will help to track the equipment (including information about when and where the equipment was manufactured, how it was transported and where it is currently located), monitor equipment performance (predict when it needs to be recycled, returned, or reuse). |
| Utility | This can be helpful in order to provide a comprehensive view of the equipment and resources involved in a circular value chain, allowing stakeholders to make data-driven decisions and optimize their operations. |

| Data Element ID | TUE_META |
|---|---|
| Origin | TUE will produce meta data descriptors to translate source data into system knowledge graphs and to conduct automated analysis on system knowledge graphs and to export data from system knowledge graphs for use cases or other demonstrators |
| Type(s) | • Raw Input data<br>• Technical data descriptors (technical meta-data)<br>• Semantic data descriptors (semantic meta-data) |
| Format(s) | • Comma Separated Values (.csv)<br>• JavaScript Object Notation (.json)<br>• Graph-based data storage (.rdf, .graphml, .gxl or other XML-based graph exchange formats)<br>• Image files (.jpg, .png, .eps)<br>• Document files (.pdf, .docx, .pptx) |
| Purpose | The data is required to construct and analyse system knowledge graphs from input event data in a repeatable and automated form. |
| Utility | The research community may use the data as a benchmark for constructing and analysing system knowledge graph. The research community may use the data to develop standards for data import, analysis, and export from system knowledge graphs. |

| Data Element ID | TUE_GRAPH |
|---|---|
| Origin | TUE will produce system knowledge graphs and related data streams describing domain knowledge and behaviour (historic, simulated, estimated) and other information for use cases or other demonstrators. |
| Type(s) | • Knowledge graph (labelled property graph)<br>• Schema of knowledge graph |

|  | • Data streams to construct knowledge graphs<br>• Data streams obtained from knowledge graphs |
|---|---|
| **Format(s)** | • Comma Separated Values (.csv)<br>• JavaScript Object Notation (.json)<br>• Neo4j database file (.db)<br>• Graph-based data storage (.rdf,.graphml, .gxl or other XML-based graph exchange formats)<br>• Image files (.jpg, .png, .eps)<br>• Document files (.pdf, .docx, .pptx) |
| **Purpose** | The data represents structural, historic, and future system knowledge and can help the development and validation of process mining techniques and other techniques for the automated construction and maintenance of digital twins. |
| **Utility** | The research community may use the data as a benchmark for developing new process mining and digital twin learning methods. The research community may use the data to develop standards for data import, analysis, and export from system knowledge graphs. |

| Data Element ID | # ERQ_ARC_DB |
|---|---|
| **Origin** | End Users |
| **Type(s)** | Entity data (presented on excel files) |
| **Format(s)** | • Comma Separated Values (.csv)<br>• Excel files (.xlsx) |
| **Purpose** | These data provide input regarding to end-users, like job role (archetype), tasks, tools and soft skills |
| **Utility** | These are used to extract information about worker occupation |

| Data Element ID | # ERQ_ARC_MCR_CL |
|---|---|
| **Origin** | End Users |
| **Type(s)** | Entity data (presented on excel files) |
| **Format(s)** | • Comma Separated Values (.csv)<br>• Excel files (.xlsx) |
| **Purpose** | These data provide input regarding to job role (archetype) and the group that they belong |
| **Utility** | These are used to extract information about job roles (archetypes) and groups mapping |

| Data Element ID | # ERQ_SFT_SKL_QUE |
|---|---|
| **Origin** | End Users |

| Type(s) | Entity data (presented on excel files) |
| --- | --- |
| Format(s) | • Comma Separated Values (.csv)<br>• Excel files (.xlsx) |
| Purpose | These data provide a list of questions related to soft skills |
| Utility | These data can be used to extract useful information related to soft skills |

| Data Element ID | # ERQ_ARC_AI |
| --- | --- |
| Origin | End Users |
| Type(s) | Entity data (presented on excel files) |
| Format(s) | • Comma Separated Values (.csv)<br>• Excel files (.xlsx) |
| Purpose | These data provide input regarding to AI components and job roles (archetypes) |
| Utility | These data can be used to extract useful information related to job descriptions (archetypes) to AI components mapping |

| Data Element ID | # ERQ_ARC_CLS_TOOL |
| --- | --- |
| Origin | End Users |
| Type(s) | Entity data (presented on excel files) |
| Format(s) | • Comma Separated Values (.csv)<br>• Excel files (.xlsx) |
| Purpose | These data provide input regarding to job roles (archetypes), tool clusters and tools |
| Utility | These data can be used to extract useful information related to job descriptions (archetypes) to tools mapping |

| Data Element ID | COR_1_Newsletter |
| --- | --- |
| Origin | Mailing List |
| Type(s) | Secondary data |
| Format(s) | csv or xlsx |
| Purpose | To send out the project's Newsletters |

| Utility | Data used to disseminate and communicate the project's results and progress to its stakeholders and target groups |
|---|---|

| Data Element ID | COR_2_SocialMedia |
|---|---|
| Origin | Contacts & Followers on Twitter & LinkedIn |
| Type(s) | Secondary data |
| Format(s) | N/A |
| Purpose | Part of the D&C activities of the project |
| Utility | The Data will also be used as KPIs of the project |

| Data Element ID | COR_3_Website |
|---|---|
| Origin | Cookies |
| Type(s) | Secondary data |
| Format(s) | N/A |
| Purpose | Part of the D&C activities of the project |
| Utility | The Data will also be used as KPIs of the project |

| Data Element ID | COR_4_EventsParticipants |
|---|---|
| Origin | Events Participants List |
| Type(s) | Secondary data |
| Format(s) | Excel files (.xlsx) |
| Purpose | Data used only for project purposes |
| Utility | N/A |

## 3   FAIR data

*It should be noted that the data management plan constitutes a living document following the guidelines of the EC, as mentioned before. Furthermore, at the moment of writing two of the project's use cases are still in prototyping phase thus the exact nature of the available data will be more clear after M12. Towards this end, the data management plan will be updated and optimised with additional and more detailed information following the advancements of the project during the complete project lifetime.*

## 3.1   Making data findable & provisions of metadata

Within the context of AUTO-TWIN, the well-established and OpenAIRE's universal repository Zenodo will be utilised as the core tool for maximizing the findability of the data in accordance with the OpenAIRE requirements.

To achieve this, the AUTO-TWIN Zenodo community has been created on Zenodo[2]. AUTO-TWIN plans to populate it and enrich it with all the datasets and deliverables which are marked as public by the consortium as well as the produced in the course of the project scientific publications. Furthermore, the means provided by Zenodo to directly link all elements of the community to the OpenAIRE community will be also leveraged.

It should be noted, as explained also in Section 3.2, that not all datasets of the project are considered as public due to several restrictions (such as privacy, security and business exploitation potential), as assessed by the consortium before they are marked as public. However, the consortium is committed on providing the maximum number of public datasets possible to the research community in accordance with industrial property rights and the rules for exploitation agreed by the consortium in the exploitation plan.

In accordance with the OpenAIRE requirements, Zenodo enables the provisioning of rich metadata that include both the Grant Number and Project Acronym on each element while unique Digital Object Identifiers (DOIs) are allocated to all the elements of the community. Furthermore, the appropriate versioning of the elements is facilitated through Zenodo tools. Finally, the consortium will apply on element a naming convention that directly associates each element with its owner and the data of day of creation.

In particular, AUTO-TWIN will apply the following metadata scheme on each element of its Zenodo community:

| Metadata ID | Short description |
|---|---|
| Digital Object Identifier | The unique digital identifier assigned to the data element |
| Community | Set by default to the AUTO-TWIN community while additional communities might be added |
| Publication date | The date of first publication of the data element |
| Title | The title of the data element |

---

| | |
|---|---|
| **Authors** | The list of authors (owners) of the data element |
| **Description** | The brief description of the data element |
| **Version** | The version number of the data element |
| **Language** | The language utilised in the data element |
| **Keywords** | A list of keywords related to the data element |
| **Access Right** | The applicable access right level of the data element |
| **License** | The applicable license of the data element |
| **Grants** | The grant information (EU, Grant Number, Project Acronym) |
| **References** | The relevant references of the data element |
| **Journal, Conference, Book, Report, Chapter** | The relevant publications and reports associated with the data element |

## 3.2 Making data openly accessible

AUTO-TWIN follows the principle ''as open as possible and as closed as necessary'' since some of the for-profit entities wish to protect their innovations, and several key artefacts that will be produced in the project will have commercial exploitation potential. Nevertheless, the consortium plans to share with the scientific community through its Zenodo community, specific subsets of the data elements which will be collected, generated or processed in the proper processed form that adheres to the desired privacy, security and commercial restrictions. AUTO-TWIN will assess all datasets and define their level of confidentiality based on the input of concerned business actors in order to keep critical business knowledge hidden from competitors and to safeguard the industrial property rights taking also into consideration the exploitation plan of the project.

In addition to the above, the consortium also aims at the usage of Open Access to scientific work towards the increase of the circulation and exploitation of the produced knowledge. To this end, AUTO-TWIN is committed to publish at least 80% of AUTO-TWIN scientific publications, in line with the policies of each partner, in Open Access form, both self-archiving (i.e. Green Open Access), or Open access publishing (i.e. Gold Open Access). Finally, AUTO-TWIN is committed to make available in a public form most significant AUTO-TWIN reports through its Zenodo community and where possible to create the adapted versions of the reports in order to properly address any restriction that prohibits their sharing.

In the following tables, the details of each data element listed in Table 1 regarding the parts that will be made publicly available are presented, providing also the necessary clarifications for any legal, contractual, IPR or other restrictions that are applied which prohibit their sharing.

| Data Element ID | # TUE_PM |
|---|---|

| Origin | TUE will compare generated from original data in the graph, enriching and exporting the final data for Process Mining |
|---|---|
| List of data element's parts that will be made openly accessible | Data that are generated from the current state of equipment |
| Utilised open repositories | TBD |
| List of data elements that will not be publicly shared | The actions of end-users will not be openly shared due to confidential reasons |
| Method / software tools required to access the data | TBD |

| Data Element ID | # SYX_GR3N_TED |
|---|---|
| Origin | SYX will enrich and store GR3N pilot user data |
| List of data element's parts that will be made openly accessible | TBD |
| Utilised open repositories | TBD |
| List of data elements that will not be publicly shared | The actions of end-users will not be openly shared due to confidential reasons |
| Method / software tools required to access the data | TBD |

| Data Element ID | # SYX_CROMA_TED |
|---|---|
| Origin | SYX will enrich and store CROMA pilot user data |
| List of data element's parts that will be made openly accessible | TBD |

| Utilised open repositories | TBD |
|---|---|
| List of data elements that will not be publicly shared | The actions of end-users will not be openly shared due to confidential reasons |
| Method / software tools required to access the data | TBD |

| Data Element ID | # SYX_LIB_TED |
|---|---|
| Origin | SYX will enrich and store Libattion pilot user data |
| List of data element's parts that will be made openly accessible | TBD |
| Utilised open repositories | TBD |
| List of data elements that will not be publicly shared | The actions of end-users will not be openly shared due to confidential reasons |
| Method / software tools required to access the data | TBD |

| Data Element ID | PMI_PM |
|---|---|
| Origin | PMI will produce process model data for use cases or other demonstrators |
| List of data element's parts that will be made openly accessible | • Process model data for Libattion's use case<br>• Process model data for GR3N's use case<br>• Process model data for Croma's use case<br>• Process model data for other demonstrators |
| Utilised open repositories | Zenodo |
| List of data elements that will not be publicly shared | Any data elements that are considered confidential by use case owners or other partners will not be publicly shared |

| Method / software tools required to access the data | No specific methods or tools are required to access the data |
| --- | --- |

| Data Element ID | # TTS_DES |
| --- | --- |
| Origin | TTS generates data from "learned" simulation |
| List of data element's parts that will be made openly accessible | Data that are generated from the current state of equipment |
| Utilised open repositories | TBD |
| List of data elements that will not be publicly shared | The actions of end-users will not be openly shared due to confidential reasons |
| Method / software tools required to access the data | TBD |

| Data Element ID | TUE_META |
| --- | --- |
| Origin | TUE will produce meta data for system knowledge graphs for use cases or other demonstrators |
| List of data element's parts that will be made openly accessible | • System knowledge graph meta-data for Libattion's use case<br>• System knowledge graph meta-data for GR3N's use case<br>• System knowledge graph meta-data for Croma's use case<br>• System knowledge graph meta-data for other demonstrators |
| Utilised open repositories | Zenodo |
| List of data elements that will not be publicly shared | Any data elements or semantic information that are considered confidential by use case owners or other partners will not be publicly shared |
| Method / software tools required to access the data | No specific methods or tools are required to access the data |

| Data Element ID | TUE_GRAPH |
| --- | --- |

| Origin | TUE will produce graph data (in graph format and related data streams) for system knowledge graphs for use cases or other demonstrators |
|---|---|
| **List of data element's parts that will be made openly accessible** | • System knowledge graph data for Libattion's use case<br>• System knowledge graph data for GR3N's use case<br>• System knowledge graph data for Croma's use case<br>• System knowledge graph data for other demonstrators |
| **Utilised open repositories** | Zenodo |
| **List of data elements that will not be publicly shared** | Any data elements or semantic information that are considered confidential by use case owners or other partners will not be publicly shared |
| **Method / software tools required to access the data** | Data will be shared in an openly accessible exchange format requiring no specific tools or methods, and in a graph database exchange format that will require a graph database system to use (e.g., neo4j). |

| Data Element ID | # ERQ_ARC_DB |
|---|---|
| **Origin** | End Users |
| **List of data element's parts that will be made openly accessible** | TBD |
| **Utilised open repositories** | TBD |
| **List of data elements that will not be publicly shared** | The actions of end-users will not be openly shared due to confidential reasons |
| **Method / software tools required to access the data** | TBD |

| Data Element ID | # ERQ_ARC_MCR_CL |
|---|---|
| **Origin** | End Users |
| **List of data element's parts that will be made openly accessible** | TBD |

| Utilised open repositories | TBD |
|---|---|
| List of data elements that will not be publicly shared | The actions of end-users will not be openly shared due to confidential reasons |
| Method / software tools required to access the data | TBD |

| Data Element ID | # ERQ_SFT_SKL_QUE |
|---|---|
| Origin | End Users |
| List of data element's parts that will be made openly accessible | TBD |
| Utilised open repositories | TBD |
| List of data elements that will not be publicly shared | The actions of end-users will not be openly shared due to confidential reasons |
| Method / software tools required to access the data | TBD |

| Data Element ID | # ERQ_ARC_AI |
|---|---|
| Origin | End Users |
| List of data element's parts that will be made openly accessible | TBD |
| Utilised open repositories | TBD |
| List of data elements that will not be publicly shared | The actions of end-users will not be openly shared due to confidential reasons |

| Method / software tools required to access the data | TBD |
|---|---|

| Data Element ID | # ERQ_ARC_CLS_TOOL |
|---|---|
| Origin | End Users |
| List of data element's parts that will be made openly accessible | TBD |
| Utilised open repositories | TBD |
| List of data elements that will not be publicly shared | The actions of end-users will not be openly shared due to confidential reasons |
| Method / software tools required to access the data | TBD |

| Data Element ID | COR_1_Newsletter |
|---|---|
| Origin | Mailing List |
| List of data element's parts that will be made openly accessible | N/A |
| Utilised open repositories | N/A |
| List of data elements that will not be publicly shared | Access to the specific data is restricted to the project's communication coordinator (COR) |
| Method / software tools required to access the data | The mailing list data is available only through the project's Mailchimp account |

| Data Element ID | COR_2_SocilaMedia |
|---|---|

| Origin | Contacts & Followers on Twitter & LinkedIn |
|---|---|
| List of data element's parts that will be made openly accessible | N/A |
| Utilised open repositories | N/A |
| List of data elements that will not be publicly shared | Access to the specific data is restricted to the project's communication coordinator (COR) |
| Method / software tools required to access the data | The data is accessible only by logging in the project's social media accounts |

| Data Element ID | COR_3_Website |
|---|---|
| Origin | Cookies - The dataset originates from the project's website (Squarespace). |
| List of data element's parts that will be made openly accessible | N/A |
| Utilised open repositories | N/A |
| List of data elements that will not be publicly shared | Access to the specific data is restricted to the project's communication coordinator (COR) |
| Method / software tools required to access the data | The data is available only through logging in to the website's CMS platform |

| Data Element ID | COR_4_EventsParticipants |
|---|---|
| Origin | Events Participants List |
| List of data element's parts that will be made openly accessible | N/A |

| Utilised open repositories | N/A |
|---|---|
| List of data elements that will not be publicly shared | Access to the specific data is restricted to the project's communication coordinator (COR). Also, access will be given to authorized people from each partner. |
| Method / software tools required to access the data | Project's repository (Nextcloud) |

## 3.3 Making data interoperable

Within the context of AUTO-TWIN, data interoperability is ensured by adopting the OpenAIRE guidelines (accessed at: https://guidelines.openaire.eu/en/latest/) which clearly define the instructions and rules that should be followed in order to provide data elements in a way that is compatible with OpenAIRE infrastructure. Since Zenodo constitutes the OpenAIRE's universal repository and it's fully compliant to the OpenAIRE's principles and guidelines, the utilised AUTO-TWIN community in Zenodo and the usage of the provided tools by Zenodo ensures the maximization of the data interoperability for the data elements of AUTO-TWIN.

The following tables provide the details for each data element listed in Table 1 with regards to the supported data exchange formats, the foreseen stakeholders of the data elements, the list of metadata vocabularies, standard or methods utilised and finally the applied data handling methods.

| Data Element ID | # TUE_PM |
|---|---|
| Origin | TUE will compare generated from original data in the graph, enriching and exporting the final data for Process Mining |
| Supported data exchange formats | <ul><li>Comma Separated Values (.csv)</li><li>XML</li><li>OCEL</li><li>Excel files (.xlsx)</li></ul> |
| Candidate stakeholders | All partners that are involved in process mining tasks |
| Metadata, schemas, ontologies utilised | N/A |
| Applied data handling methods | TBD |

| Data Element ID | # SYX_GR3N_TED |
|---|---|
| Origin | SYX will enrich and store GR3N pilot user data |
| Supported data exchange formats | <ul><li>Comma Separated Values (.csv)</li><li>JavaScript Object Notation (.json)</li><li>Apache Cassandra database files (.db)</li><li>MongoDB database files (.bson)</li></ul> |

| | |
|---|---|
| | • Image files (.jpg, .png, .eps)<br>• Document files (.pdf, .docx, .pptx) |
| **Candidate stakeholders** | All technical partners |
| **Metadata, schemas, ontologies utilised** | TBD |
| **Applied data handling methods** | TBD |

| | |
|---|---|
| **Data Element ID** | **# SYX_CROMA_TED** |
| **Origin** | SYX will enrich and store CROMA pilot user data |
| **Supported data exchange formats** | • Comma Separated Values (.csv)<br>• JavaScript Object Notation (.json)<br>• Apache Cassandra database files (.db)<br>• MongoDB database files (.bson)<br>• Image files (.jpg, .png, .eps)<br>• Document files (.pdf, .docx, .pptx) |
| **Candidate stakeholders** | All technical partners |
| **Metadata, schemas, ontologies utilised** | TBD |
| **Applied data handling methods** | TBD |

| | |
|---|---|
| **Data Element ID** | **# SYX_LIB_TED** |
| **Origin** | SYX will enrich and store Libattion pilot user data |
| **Supported data exchange formats** | • Comma Separated Values (.csv)<br>• JavaScript Object Notation (.json)<br>• Apache Cassandra database files (.db)<br>• MongoDB database files (.bson)<br>• Image files (.jpg, .png, .eps)<br>• Document files (.pdf, .docx, .pptx) |
| **Candidate stakeholders** | All technical partners |
| **Metadata, schemas, ontologies utilised** | TBD |
| **Applied data handling methods** | TBD |

| Data Element ID | PMI_PM |
|---|---|
| Origin | PMI will produce process model data for use cases or other demonstrators |
| Supported data exchange formats | • Comma Separated Values (.csv)<br>• JavaScript Object Notation (.json)<br>• Neo4j database files (.db)<br>• Image files (.jpg, .png, .eps)<br>• Document files (.pdf, .docx, .pptx) |
| Candidate stakeholders | All partners that are involved in the generation of process models |
| Metadata, schemas, ontologies utilised | No specific metadata, schemas or ontologies will be utilised |
| Applied data handling methods | TBD |

| Data Element ID | # TTS_DES |
|---|---|
| Origin | TTS generates data from "learned" simulation |
| Supported data exchange formats | • Comma Separated Values (.csv)<br>• XML<br>• JSON<br>• Excel files (.xlsx) |
| Candidate stakeholders | All partners that are involved in the construction of knowledge graphs (GraphDB) |
| Metadata, schemas, ontologies utilised | N/A |
| Applied data handling methods | TBD |

| Data Element ID | TUE_META |
|---|---|
| Origin | TUE will produce meta data for system knowledge graphs for use cases or other demonstrators |
| Supported data exchange formats | • Comma Separated Values (.csv)<br>• JavaScript Object Notation (.json)<br>• Graph-based data storage (.rdf, .graphml, .gxl or other XML-based graph exchange formats)<br>• Image files (.jpg, .png, .eps)<br>• Document files (.pdf, .docx, .pptx) |

| Candidate stakeholders | All partners that are involved in the construction of system knowledge graphs. |
|---|---|
| Metadata, schemas, ontologies utilised | No specific metadata, schemas or ontologies will be utilised |
| Applied data handling methods | TBD |

| Data Element ID | TUE_GRAPH |
|---|---|
| Origin | TUE will produce graph data (in graph format and related data streams) for system knowledge graphs for use cases or other demonstrators |
| Supported data exchange formats | • Comma Separated Values (.csv)<br>• JavaScript Object Notation (.json)<br>• Neo4j database file (.db)<br>• Graph-based data storage (.rdf, .graphml, .gxl or other XML-based graph exchange formats) |
| Candidate stakeholders | All partners that are involved in the construction of system knowledge graphs. |
| Metadata, schemas, ontologies utilised | The data will be produced based on the TUE_META meta data. |
| Applied data handling methods | TBD |

| Data Element ID | # ERQ_ARC_DB |
|---|---|
| Origin | End Users |
| Supported data exchange formats | • Comma Separated Values (.csv)<br>• Excel files (.xlsx) |
| Candidate stakeholders | TBD |
| Metadata, schemas, ontologies utilised | TBD |
| Applied data handling methods | TBD |

| Data Element ID | # ERQ_ARC_MCR_CL |
|---|---|

| Origin | End Users |
|---|---|
| **Supported data exchange formats** | <ul><li>Comma Separated Values (.csv)</li><li>Excel files (.xlsx)</li></ul> |
| **Candidate stakeholders** | TBD |
| **Metadata, schemas, ontologies utilised** | TBD |
| **Applied data handling methods** | TBD |

| Data Element ID | # ERQ_SFT_SKL_QUE |
|---|---|
| **Origin** | End Users |
| **Supported data exchange formats** | <ul><li>Comma Separated Values (.csv)</li><li>Excel files (.xlsx)</li></ul> |
| **Candidate stakeholders** | TBD |
| **Metadata, schemas, ontologies utilised** | TBD |
| **Applied data handling methods** | TBD |

| Data Element ID | # ERQ_ARC_AI |
|---|---|
| **Origin** | End Users |
| **Supported data exchange formats** | <ul><li>Comma Separated Values (.csv)</li><li>Excel files (.xlsx)</li></ul> |
| **Candidate stakeholders** | TBD |
| **Metadata, schemas, ontologies utilised** | TBD |
| **Applied data handling methods** | TBD |

| Data Element ID | # ERQ_ARC_CLS_TOOL |
|---|---|
| Origin | End Users |
| Supported data exchange formats | • Comma Separated Values (.csv)<br>• Excel files (.xlsx) |
| Candidate stakeholders | TBD |
| Metadata, schemas, ontologies utilised | TBD |
| Applied data handling methods | TBD |

| Data Element ID | COR_1_Newsletter |
|---|---|
| Origin | Mailing List |
| Supported data exchange formats | csv / xlsx |
| Candidate stakeholders | It is restricted to the communication coordinator of the project (COR) for dissemination purposes |
| Metadata, schemas, ontologies utilised | N/A |
| Applied data handling methods | Encrypted storage |

| Data Element ID | COR_2_SocilaMedia |
|---|---|
| Origin | Contacts & Followers on Twitter & LinkedIn |
| Supported data exchange formats | N/A |
| Candidate stakeholders | It is restricted to the communication coordinator of the project (COR) for dissemination purposes |
| Metadata, schemas, ontologies utilised | N/A |
| Applied data handling methods | Encrypted storage |

| Data Element ID | COR_3_Website |
|---|---|
| Origin | Website Cookies |
| Supported data exchange formats | N/A |
| Candidate stakeholders | It is restricted to the communication coordinator of the project (COR) for dissemination purposes |
| Metadata, schemas, ontologies utilised | N/A |
| Applied data handling methods | Encrypted storage |

| Data Element ID | COR_4_EventsParticipants |
|---|---|
| Origin | Events Participants List |
| Supported data exchange formats | xlsx |
| Candidate stakeholders | All the partners that are interested to expand their contact lists |
| Metadata, schemas, ontologies utilised | N/A |
| Applied data handling methods | Encrypted storage |

## 3.4  Increase data re-use (licensing)

The consortium will leverage the tools provided by Zenodo for the enforcement of policies related to the access to the data elements as well as any embargo period that can be applied. For each data element a specific license will be applied ranging from MIT license, to CC BY 4.0 License and more depending on the nature and specifics of each data element.

Zenodo also guarantees the long data availability of the published data elements as it ensures a lifetime of the repository for at least the next 20 years. In addition to this, Zenodo offers an advanced and sustainable archiving process which ensures the reusability of the data elements. Furthermore, Zenodo facilitates the secure storage of the data elements on its repositories with state-of-the-art security mechanisms as well as a sophisticated data lineage process for the data elements and their metadata which provides tracing capabilities and quality assurance features.

Besides Zenodo, public and confidential data elements will be stored and managed, archived and deleted on the on-premises systems of the participants of the consortium who will also apply the best

practises in the areas of security, access control and auditing as elaborated in Section 4 of the current deliverable.

In the following tables, the details of each data element listed in Table 1 with regards to the applied methods for data release, data archiving and the applied license are presented.

| Data Element ID | # TUE_PM |
|---|---|
| Origin | TUE will compare generated from original data in the graph, enriching and exporting the final data for Process Mining |
| Applied method for data release | All data generated during the project will be encrypted, archived and stored in a secure and trusted storage. The storage of the generated data is ensured for 3 years. Backup of the stored data will be also maintained in a separated secure environment. |
| Applied method for data archiving | The data will be discarded 3 years after the end of the AUTO-TWIN project |
| Licensing information | MIT License will be applied |

| Data Element ID | # SYX_GR3N_TED |
|---|---|
| Origin | SYX will enrich and store GR3N pilot user data |
| Applied method for data release | All data generated during the project will be encrypted, archived and stored in a secure and trusted storage. The storage of the generated data is ensured for 3 years. Backup of the stored data will be also maintained in a separated secure environment |
| Applied method for data archiving | The data will be discarded 3 years after the end of the AUTO-TWIN project |
| Licensing information | MIT License will be applied |

| Data Element ID | # SYX_CROMA_TED |
|---|---|
| Origin | SYX will enrich and store CROMA pilot user data |
| Applied method for data release | All data generated during the project will be encrypted, archived and stored in a secure and trusted storage. The storage of the generated data is ensured for 3 years. Backup of the stored data will be also maintained in a separated secure environment |
| Applied method for data archiving | The data will be discarded 3 years after the end of the AUTO-TWIN project |
| Licensing information | MIT License will be applied |

| Data Element ID | # SYX_LIB_TED |
|---|---|
| Origin | SYX will enrich and store LIB pilot user data |
| Applied method for data release | All data generated during the project will be encrypted, archived and stored in a secure and trusted storage. The storage of the generated data is ensured for 3 years. Backup of the stored data will be also maintained in a separated secure environment |
| Applied method for data archiving | The data will be discarded 3 years after the end of the AUTO-TWIN project |
| Licensing information | MIT License will be applied |

| Data Element ID | PMI_PM |
|---|---|
| Origin | PMI will produce process model data for use cases or other demonstrators |
| Applied method for data release | TBD |
| Applied method for data archiving | TBD |
| Licensing information | CC BY 4.0 License will be applied |

| Data Element ID | # TTS_DES |
|---|---|
| Origin | TTS generates data from "learned" simulation |
| Applied method for data release | All data generated during the project will be encrypted, archived and stored in a secure and trusted storage. The storage of the generated data is ensured for 3 years. Backup of the stored data will be also maintained in a separated secure environment… |
| Applied method for data archiving | The data will be discarded 3 years after the end of the AUTO-TWIN project |
| Licensing information | MIT License will be applied |

| Data Element ID | TUE_META |
|---|---|
| Origin | TUE will produce meta data for system knowledge graphs for use cases or other demonstrators |

| | |
|---|---|
| **Applied method for data release** | TBD |
| **Applied method for data archiving** | TBD |
| **Licensing information** | CC BY 4.0 License will be applied |

| Data Element ID | TUE_GRAPH |
|---|---|
| **Origin** | TUE will produce graph data (in graph format and related data streams) for system knowledge graphs for use cases or other demonstrators |
| **Applied method for data release** | TBD |
| **Applied method for data archiving** | TBD |
| **Licensing information** | CC BY 4.0 License will be applied |

| Data Element ID | # ERQ_ARC_DB |
|---|---|
| **Origin** | End Users |
| **Applied method for data release** | All data generated during the project will be encrypted, archived and stored in a secure and trusted storage. The storage of the generated data is ensured for 3 years. Backup of the stored data will be also maintained in a separated secure environment |
| **Applied method for data archiving** | TBD |
| **Licensing information** | MIT License will be applied |

| Data Element ID | # ERQ_ARC_MCR_CL |
|---|---|
| **Origin** | End Users |
| **Applied method for data release** | All data generated during the project will be encrypted, archived and stored in a secure and trusted storage. The storage of the generated data is ensured for 3 years. Backup of the stored data will be also maintained in a separated secure environment |
| **Applied method for data archiving** | TBD |

| Licensing information | MIT License will be applied |
|---|---|

| Data Element ID | # ERQ_SFT_SKL_QUE |
|---|---|
| Origin | End Users |
| Applied method for data release | All data generated during the project will be encrypted, archived and stored in a secure and trusted storage. The storage of the generated data is ensured for 3 years. Backup of the stored data will be also maintained in a separated secure environment |
| Applied method for data archiving | TBD |
| Licensing information | MIT License will be applied |

| Data Element ID | # ERQ_ARC_AI |
|---|---|
| Origin | End Users |
| Applied method for data release | All data generated during the project will be encrypted, archived and stored in a secure and trusted storage. The storage of the generated data is ensured for 3 years. Backup of the stored data will be also maintained in a separated secure environment |
| Applied method for data archiving | TBD |
| Licensing information | MIT License will be applied |

| Data Element ID | # ERQ_ARC_CLS_TOOL |
|---|---|
| Origin | End Users |
| Applied method for data release | All data generated during the project will be encrypted, archived and stored in a secure and trusted storage. The storage of the generated data is ensured for 3 years. Backup of the stored data will be also maintained in a separated secure environment |
| Applied method for data archiving | TBD |
| Licensing information | MIT License will be applied |

| Data Element ID | COR_1_Newsletter |
|---|---|
| **Origin** | Mailing List |
| **Applied method for data release** | This particular dataset will be encrypted and stored in a secure location throughout the project. Both the data and the backups will be held in a different secure cloud owned by COR. |
| **Applied method for data archiving** | The data will be discarded 3 years after the end of the AUTO-TWIN project |
| **Licensing information** | Access restricted to the communication officer of the project (COR) |

| Data Element ID | COR_2_SocialMedia |
|---|---|
| **Origin** | Contacts & Followers on Twitter & LinkedIn |
| **Applied method for data release** | N/A |
| **Applied method for data archiving** | No use beyond the project |
| **Licensing information** | Access restricted to the communication officer of the project (COR) |

| Data Element ID | COR_3_Website |
|---|---|
| **Origin** | Cookies |
| **Applied method for data release** | This particular dataset will be encrypted and stored in a secure location throughout the project. Both data and the backups will be held in a different secure cloud owned by COR. |
| **Applied method for data archiving** | The data will be discarded 3 years after the end of the AUTO-TWIN project |
| **Licensing information** | Access restricted to the communication officer of the project (COR) |

| Data Element ID | COR_4_EventsParticipants |
|---|---|
| **Origin** | Events Participant Lists |
| **Applied method for data release** | This particular dataset will be encrypted and stored in a secure location throughout the project. Both data and the backups will be held in a different secure cloud owned by COR. |

| | |
|---|---|
| **Applied method for data archiving** | The data will be discarded 3 years after the end of the AUTO-TWIN project |
| **Licensing information** | Authorized partners only & the communication officer (COR) |

## 4    Data Management allocation of resources

*It should be noted that the data management plan constitutes a living document following the guidelines of the EC, as mentioned before. Furthermore, at the moment of writing two of the project's use cases are still in prototyping phase thus the exact nature of the available data will be more clear after M12. Towards this end, the data management plan will be updated and optimised with additional and more detailed information following the advancements of the project during the complete project lifetime.*

As documented in the previous sections, AUTO-TWIN has selected the OpenAIRE's universal repository Zenodo as the core tool for managing and boosting the FAIR aspects of the project's open data. Zenodo is a free-of-charge repository offering out of the box the maximum compliance with the FAIR principles with services that facilitate the maximization of the findability, open access, interoperability, re-use, safe storage, versioning, archiving and long preservation of data.

COR is responsible for the handling of the project's Zenodo community on behalf of AUTO-TWIN in collaboration with UBI, who is the leader of T7.5, and PMI, who is the leader of WP7. In addition to this, all data management aspects and issues will be resolved by the UBI in collaboration with the PMI. Finally, each partner is responsible for the data they collect, generate and process as elaborated also in Section 4.

# 5   Data Security

*It should be noted that the data management plan constitutes a living document following the guidelines of the EC, as mentioned before. Furthermore, at the moment of writing two of the project's use cases are still in prototyping phase thus the exact nature of the available data will be more clear after M12. Towards this end, the data management plan will be updated and optimised with additional and more detailed information following the advancements of the project during the complete project lifetime.*

With regards to Zenodo, the offered service is safeguarded by the data security policies applied in CERN Data Centers which ensure maximum security compliance. In particular, data are stored only within their data centers while daily backups of data and metadata are performed. In the background, integrity checks such as checksum validation are performed while Zenodo guarantees that in case of the service closure or discontinuation all data and metadata will be transferred to other repositories which provide the same principles.

With regards to the data that will be collected, processed and stored by the consortium partners in the course of the project, each partner will be responsible for the data that they collect, process and store. Each partner will ensure the compliance with the regulations of the countries where the partner resides. Partners are committed to use repositories that apply the highest security standards where strict policies are enforced and verified through the appropriate risk assessment processes. The repositories will be certified and only accessible by authorised personnel ensuring that best practises in data security are applied at all times with various provisions such as encryption tools, secure transfer protocols (https for web and sftp for local data transfer) depending on the needs of the project.

With regards to data retention, which is the responsibility of the respective partner, the details are provided in the relevant tables of section 3.4 and the partners will ensure data will be destroyed as soon as the defined retention period ends. As a principle, all data residing on the premises of the partners will be discarded automatically in a maximum of three years after the end of the project and any possible personal data will be destroyed once the project has been completed unless data owners have provided an explicit consent to keep them.

With regards to data backup, it is the responsibility of the respective partner to ensure that regular back-ups are performed and securely stored during the whole project period with the appropriate storage medium in the appropriate storage locations.

# 6   Ethical aspects

*It should be noted that the data management plan constitutes a living document following the guidelines of the EC, as mentioned before. Furthermore, at the moment of writing two of the project's use cases are still in prototyping phase thus the exact nature of the available data will be more clear after M12. Towards this end, the data management plan will be updated and optimised with additional and more detailed information following the advancements of the project during the complete project lifetime.*

In accordance with the DoA, Deliverable D8.1 "OEI - Requirement No. 1" has been delivered on M3 with the scope of providing a detailed Ethics Summary Report which describes the activities that the consortium will put in place to grant the maximum protection of personal data and to fulfil the EU ethics requirements. As, the AUTO-TWIN consortium is fully aware of the ethical implications of the performed research study, deliverable D8.1 defines the requirements which are met in order to conform to national and international ethical guidelines. Hence, the details of parts of the data elements which require ethics approval, inform consent and long-term preservation questionnaires were already elaborated in the aforementioned deliverable.

The scope of the current section is to provide the supplementary documentation of the details of all data elements for the ethical perspective and to constantly monitor and update their ethical aspect details within the course of the project.

| Data Element ID | # TUE_PM |
|---|---|
| Origin | TUE will compare generated from original data in the graph, enriching and exporting the final data for Process Mining |
| Personal Data included | TBD |
| Description of data that may require ethical approval | TBD |
| Applicable standards, policies, legislation etc | GDPR |
| Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc | TBD |

| Data Element ID | # SYX_GR3N_TED |
|---|---|

| Origin | SYX will enrich and store GR3N pilot user data |
|---|---|
| **Personal Data included** | TBD |
| **Description of data that may require ethical approval** | TBD |
| **Applicable standards, policies, legislation etc** | GDPR |
| **Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc** | TBD |

| **Data Element ID** | **# SYX_CROMA_TED** |
|---|---|
| Origin | SYX will enrich and store CROMA pilot user data |
| **Personal Data included** | TBD |
| **Description of data that may require ethical approval** | TBD |
| **Applicable standards, policies, legislation etc** | GDPR |
| **Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc** | TBD |

| **Data Element ID** | **# SYX_LIB_TED** |
|---|---|
| Origin | SYX will enrich and store Libattion pilot user data |

| Personal Data included | TBD |
|---|---|
| Description of data that may require ethical approval | TBD |
| Applicable standards, policies, legislation etc | GDPR |
| Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc | TBD |

| Data Element ID | PMI_PM |
|---|---|
| Origin | PMI will produce process model data for use cases or other demonstrators |
| Personal Data included | No |
| Description of data that may require ethical approval | N/A |
| Applicable standards, policies, legislation etc | N/A |
| Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc | N/A |

| Data Element ID | # TTS_DES |
|---|---|
| Origin | TTS generates data from "learned" simulation |
| Personal Data included | No |

| Description of data that may require ethical approval | N/A |
|---|---|
| Applicable standards, policies, legislation etc | N/A |
| Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc | N/A |

| Data Element ID | TUE_META |
|---|---|
| Origin | TUE will produce meta data for system knowledge graphs for use cases or other demonstrators |
| Personal Data included | No |
| Description of data that may require ethical approval | N/A |
| Applicable standards, policies, legislation etc | N/A |
| Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc | N/A |

| Data Element ID | TUE_GRAPH |
|---|---|
| Origin | TUE will produce graph data (in graph format and related data streams) for system knowledge graphs for use cases or other demonstrators |
| Personal Data included | No (not known or expected at this stage). |
| Description of data that may require ethical approval | No (not known or expected at this stage). |

| | Assumption is that data provided to TUE for constructing system knowledge graphs has already been assessed for personal data and handled accordingly (by omission, anonymization, or pseudonymization on accordance with standards, policies, and legislation of the data providers). |
|---|---|
| **Applicable standards, policies, legislation etc** | |
| **Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc** | TBD |

| **Data Element ID** | **# ERQ_ARC_DB** |
|---|---|
| **Origin** | End Users |
| **Personal Data included** | No |
| **Description of data that may require ethical approval** | Authorization to save questionnaire response data |
| **Applicable standards, policies, legislation etc** | None |
| **Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc** | Deliverable D8.1 "OEI - Requirement No. 1" |

| **Data Element ID** | **# ERQ_ARC_MCR_CL** |
|---|---|
| **Origin** | End Users |
| **Personal Data included** | No |
| **Description of data that may require ethical approval** | Authorization to save questionnaire response data |

| Applicable standards, policies, legislation etc | GDPR |
|---|---|
| Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc | Deliverable D8.1 "OEI - Requirement No. 1" |

| Data Element ID | # ERQ_SFT_SKL_QUE |
|---|---|
| Origin | End Users |
| Personal Data included | No |
| Description of data that may require ethical approval | Authorization to save questionnaire response data |
| Applicable standards, policies, legislation etc | GDPR |
| Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc | Deliverable D8.1 "OEI - Requirement No. 1" |

| Data Element ID | # ERQ_ARC_AI |
|---|---|
| Origin | End Users |
| Personal Data included | No |
| Description of data that may require ethical approval | Authorization to save questionnaire response data |
| Applicable standards, policies, legislation etc | GDPR |

| Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc | Deliverable D8.1 "OEI - Requirement No. 1" |
|---|---|

| Data Element ID | # ERQ_ARC_CLS_TOOL |
|---|---|
| Origin | End Users |
| Personal Data included | No |
| Description of data that may require ethical approval | Authorization to save questionnaire response data |
| Applicable standards, policies, legislation etc | GDPR |
| Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc | Deliverable D8.1 "OEI - Requirement No. 1" |

| Data Element ID | COR_1_Newsletter |
|---|---|
| Origin | Mailing List |
| Personal Data included | Yes |
| Description of data that may require ethical approval | Name, Surname, E-mail, Occupation, Field of Expertise |
| Applicable standards, policies, legislation etc | GDPR, AUTO-TWIN Privacy policy |
| Link to related docs - Ethics approval, consents, | AUTO-TWIN Privacy Policy available at: https://www.auto-twin-project.eu/s/AUTO-TWIN-Privacy-Policy.pdf |

| Data Element ID | COR_2_SocialMedia |
|---|---|
| **Origin** | Social Media Contacts (Twitter & LinkedIn) |
| **Personal Data included** | Yes |
| **Description of data that may require ethical approval** | The dataset originates from the project's social media accounts (Twitter & LinkedIn) – Name, Surname, Location*, Date of Birth*[3], Email |
| **Applicable standards, policies, legislation etc** | GDPR, AUTO-TWIN Privacy policy, LinkedIn and Twitter Terms of Service |
| **Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc** | N/A |

| Data Element ID | COR_3_Website |
|---|---|
| **Origin** | Website Cookies |
| **Personal Data included** | Yes |
| **Description of data that may require ethical approval** | IP, Location data - The dataset originates from the project's website (Squarespace) |
| **Applicable standards, policies, legislation etc** | GDPR, AUTO-TWIN Privacy policy |
| **Link to related docs - Ethics approval, consents,** | AUTO-TWIN Privacy Policy available at: https://www.auto-twin-project.eu/s/AUTO-TWIN-Privacy-Policy.pdf |

[3] *In case the account holder has provided the information to the platform and has made it available*

| questionnaires, long-term preservation etc | |
|---|---|
| | 49 |

| Data Element ID | COR_4_EventsParticipants |
|---|---|
| Origin | Participants Lists |
| Personal Data included | Yes |
| Description of data that may require ethical approval | Name, Surname, Email, Signature or any other contact details of the events' participants. |
| Applicable standards, policies, legislation etc | GDPR, AUTO-TWIN Privacy policy |
| Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc | AUTO-TWIN Privacy Policy available at:<br><br>https://www.auto-twin-project.eu/s/AUTO-TWIN-Privacy-Policy.pdf |

# 7   Conclusions

The purpose of deliverable "D7.1 - "Data Management Plan" was to report the preliminary outcomes of the work performed within the context of Task 7.5 "Data protection and data management plan" of WP7. To this end, the deliverable documented the initial version of the AUTO-TWIN Data Management Plan by elaborating on the main principles and guidelines of the applied data management policy in AUTO-TWIN.

Towards this end, the document presented the details of the AUTO-TWIN data collection and generation process and its relation to the objectives of the project by providing the initial list of data elements which will be leveraged towards the realisation of these objectives along with their specifications. In addition to this, the document presented applied principles and guidelines on each data element which maximise their compliance with the FAIR (findable, accessible, interoperable and re-usable) principles. Furthermore, the document documented the AUTO-TWIN strategy for the data management allocation of resources while also providing a detailed description of the adopted data security strategy. Finally, the deliverable elaborated on the ethical aspects of each data element by presenting additional details of each data element supplementing the information included in deliverable D8.1.

As it's clearly stated across the document, this deliverable constitutes a living document, which will be updated and optimised with additional and more detailed information following the advancements of the project during the complete project lifetime. The upcoming versions of the deliverable will be reported on M18, M27 and M36 in accordance with the AUTO-TWIN Description of Action.