DATA MANAGEMENT PLAN

Deliverable D9.5

The TARGET-X project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No: 101096614

DATA MANAGEMENT PLAN	
GRANT AGREEMENT	101096614
PROJECT TITLE	Trial Platform foR 5G EvoluTion – Cross-Industry On Large Scale
PROJECT ACRONYM	TARGET-X
PROJECT WEBSITE	www.target-x.eu
PROJECT IDENTIFIER	https://doi.org/10.3030/101096614
PROGRAMME	HORIZON-JU-SNS-2022-STREAM-D-01-01 — SNS Large Scale Trials and Pilots (LST&Ps) with Verticals
PROJECT START	01-01-2023
DURATION	30 Months
DELIVERABLE TYPE	Deliverable
CONTRIBUTING WORK PACKAGES	WP9
DISSEMINATION LEVEL	Public
DUE DATE	M6
ACTUAL SUBMISSION DATE	M6
RESPONSIBLE ORGANIZATION	Fraunhofer IPT
EDITOR(S)	Janina Knußmann
VERSION	1.0
STATUS:	Final
SHORT ABSTRACT	Data management plan of the TARGET-X project
KEYWORDS	Data management plan, DMP, FAIR, 5G, 6G, trial sites
CONTRIBUTOR(S)	Bart Mellaerts, Maximilian Brochhaus, Marit Zöcklein, Lukas Schäper, Lucas Manassés, Manuel Pitz, Jad Nasreddine

Disclaimer

Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the other granting authorities. Neither the European Union nor the granting authority can be held responsible for them.

Executive Summary

The TARGET-X Data Management Plan (DMP) deliverable shows how the consortium partners generated and/or collected data and scientific publications are to be stored, maintained, and disseminated. The DMP serves as a guiding document for data management and is a living document that the project partners will update to reflect the changes in the generated/collected data.

The DMP focuses on how the project will provide free online access to scientific information, i.e., peer-reviewed research articles and data sets. In this context, different platforms are suggested to store and disseminate the data generated/collected such as Zenodo and GitLab. During the project lifetime, the consortium will choose the most appropriate platform in terms of fairness and reachability.

Table of Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

CORDIS	Community Research and Development Information Service
DMP	Data management plan
DoA	Description of Action
DOI	Digital object identifier
FAIR	Findable, accessible, interoperable, reusable
FSTP	Financial support for third parties
GDPR	General Data Protection Regulation
ICT	Information and communications technology
ORCID	Open Researcher and Contributor ID
SNS JU	Smart Networks and Services Joint Undertaking

1 Introduction

The purpose of the data management plan (DMP) at hand is to provide an overview of the principles, procedures and processes for data acquisition, preparation, pre-processing, analysis, storage, and distribution withing the TARGETX project. The described principles, procedures and processes will be distributed among all project members to ensure that all data from the project is appropriately dealt with during and after the project's runtime.

The document is structured as follows: Chapter 1 introduces the DMP, describes the followed methodology, further procedures for data management and used resources/infrastructures. Chapter 2 contains the data summary detailing the purpose of the data acquisition and usage, the structure of used data models (types and formats of data as well as sizes of data sets) and reusability of data sets for other purposes. Chapter 3 describes how resources will be employed to make TARGET-X data findable, accessible, interoperable and reusable (FAIR). The final chapter contains a summary of the DMP.

The TARGET-X project consortium brings together ICT as well as industry and research experts from four verticals: manufacturing, energy, construction and automotive. Therefore, the data that will be generated over the course of the project originates from different application domains which makes it more challenging to handle all data equally. For this reason, the DMP is created as a living document that can be adapted over the course of the project providing rules and guidelines to the project members for handling the different types of data.

One of the overarching objectives of TARGET-X is to make newly generated knowledge accessible and usable for the research community. The acquisition, preparation, pre-processing, analysis, storage, and distribution sustainable provision of research data will therefore be carried out according to recognized standards to make all data associated with the project FAIR. Legal and ethical obligations will be considered and characteristics like different cultural backgrounds of people involved will be taken into account in the context of data management. Furthermore, TARGET-X and the project members will observe ethical, data protection and copyright or confidentiality concerns in research data management. The examination of research data in the sense of the Employee Invention Act remains unaffected by this. In the event of a transfer of subsequent use or publication rights, care shall be taken to ensure that the data remain freely available for scientific purposes. Project leaders and independent researchers are generally responsible for the research data management of their research projects. They are obliged to ensure compliance with good scientific practice and subject standards. All people working on a research project are responsible for the correctness of the data they collect and for compliance with the regulations made.

If successfully implemented, the DMP at hand will contribute significantly to one of the key motivating factors of the project which is to establish transparency and traceability of research data. In this way, other researchers can make use of the data generated over the course of TARGETX and extend the scope of the project and its results further.

1.1 Methodology for creating the data management plan

The TARGET-X data management plan is based on the Horizon Europe Template [1]. In the first step, the data that will be expectedly created in the project such as documents and data sets have been

collected. The overview and list of data sets, which is provided in the data summary section, is not final but used as a starting point for understanding the data created in the project.

The data sets and their requirements for data management have been investigated using the ARGOS tool [2] for describing data sets and writing data management plans, also following the Horizon Europe Template there. The results of the data set descriptions have been condensed and refined into this data management plan. Further, joint approaches to identifiers, meta data and repositories have been deducted. These considerations are guided by the FAIR principles [3].

1.2 Further procedures for data management

The following chapter outlines the additional procedures and guidelines for effective data management throughout the life cycle of the TARGET-X project. These procedures ensure the integrity, accessibility, security, and usability of the data that will be collected and generated:

- Data organization and documentation:
  - Establish a consistent and standardized naming convention for files and folders to facilitate easy identification and retrieval of data.
  - Create a detailed data dictionary or metadata documentation describing the data elements, their definitions, and any associated codes or variables.
  - Maintain an inventory of all data sources, including their location, format, and access rights.
  - Clearly document data collection procedures, including protocols, equipment used, and any adjustments made during the process.
- Data storage and backup:
  - Identify appropriate data storage solutions based on the nature and volume of the data.
  - Utilize secure and reliable storage systems, such as cloud-based platforms or local servers, with sufficient capacity to accommodate data growth.
  - Implement regular data backups to prevent data loss due to hardware failures, accidental deletion, or security breaches.
  - Periodically test the data restoration process to ensure the effectiveness of backup strategies.
- Data security and access control:
  - Establish access control policies to restrict data access to authorized personnel only.
  - Implement strong user authentication mechanisms, such as username and password combinations or two-factor authentication.
  - Encrypt sensitive data during transmission and storage to protect against unauthorized access or data breaches.
  - Regularly update software and systems with security patches to address vulnerabilities and reduce the risk of cyberattacks.
- Data sharing and collaboration:
  - Define data sharing policies and procedures, specifying the conditions under which data can be shared, the appropriate data formats, and any restrictions or licenses associated with the data.
  - Utilize secure data sharing platforms or repositories suitable for the target groups that support version control, collaboration, and access management.

- Ensure compliance with legal and ethical obligations regarding data sharing, including obtaining informed consent and anonymizing personal or sensitive data when necessary.
- Establish clear communication channels and guidelines for data collaboration among project team members, ensuring proper attribution and acknowledgment of data contributions.

- Data preservation and archiving:
  - Develop a data retention schedule outlining the duration for which data will be stored and the criteria for data removal or archiving.
  - Identify appropriate long-term data archiving solutions that ensure data preservation, such as digital repositories or data centers.
  - Document the procedures for data transfer to long-term storage, including file formats, metadata preservation, and any necessary data transformations.
  - Comply with legal and regulatory requirements for data retention and archiving, considering any specific obligations related to sensitive data or intellectual property.
- Data management training and support:
  - Provide training and resources to project personnel on data management best practices, including data handling, documentation, and security protocols.
  - Offer technical support and guidance to team members to address any data management challenges or issues that may arise during the project.
  - Regularly review and update data management procedures and provide ongoing education to ensure adherence to evolving data management standards and policies.

By implementing these further procedures for data management, it is possible to maintain the integrity, security, and accessibility of the project's data throughout its life cycle. These measures promote transparency, reproducibility, and compliance with legal and ethical requirements, ultimately enhancing the value and impact of the research or project outcomes.

Overall, a consultancy of RWTH Aachen University internal research data management team and the German national research data infrastructure project for engineering sciences (NFDI4Ing) [4] is being conducted to plan and execute the data management plan.

1.3 Allocation of resources

The TARGET-X project will use free software and platforms to store the collected data in the project and make it available and FAIR. More specifically, the consortium will use the Zenodo repository and Gitlab for publishing the data. The use of such open and free repositories will allow high visibility of the data that will be available for long-term. Although the platforms are free, the TARGET-X partners will dedicate some efforts to managing repositories and uploading data. The costs for making data and other research outputs FAIR come from storage, archiving, security and indirect costs. Most of the costs are kept low by using a joint institutional infrastructure at the respective partner. Furthermore, legal costs are also expected to ensure that the collected data are compliant with legal rules.

1.4 Responsibility

Fraunhofer IPT will assume responsibility for data management in TARGET-X. The institute is also leading and coordinating the project and will ensure compliance with the guidelines, principles and instructions for action described in the DMP. Fraunhofer IPT will also assume responsibility for maintenance, updating and possible extension of the DMP. In this way, the acquisition, processing, storage, and use of research data in TARGET-X is according to jointly agreed upon guidelines is ensured.

2 Data Summary

The DMP addresses the effective management and use of data in the context of the TARGET-X project. This chapter - "Data Summary" - provides an overview of the various aspects of data related to the project. Different aspects of data handling like the use of research data and ethical issues are considered. Careful and systematic handling of data is critical to ensure the quality and sustainability of the project. This chapter serves as a basic guideline for handling data during the project and sets the framework for the later detailed description of the individual data management aspects.

2.1 Data Classification

Figure 1 depicts four categories of data that are also distinguished by three levels of confidentiality. The first category is research data which contains e.g., measurement data from the different verticals. In TARGET-X the measurement data can originate from both network performance measurements and measurements that are conducted according to the use cases implemented in each vertical. Research data is usually raw data that has not been processed or analyzed yet. This can be, for example, the consumption rate of electrical energy of a milling process that is monitored by an integrated and wirelessly connected measurement device. The second category is analyzed research data. Data from the first category is employed to generate the data from this category through data analysis and evaluation. Through analysis, information is gathered from the (raw) research data. The third data category contains information material, public reports, and deliverables as well as scientific publications that are generated and published over the course of the project. The final category is titled project data and contains internal data on the project like e.g., agreements among the project consortium and financial statements.

The confidentiality levels are structured as follows. The bottom level (partner/subcontractor) contains highly confidential data that remains with the partner or with the project members who acquired the data and is not shared with other members of the consortium. The middle level (consortium/EC) contains data that can be shared among the consortium as well as European Commission Services. The top layer (public) contains data that is not subject to any confidentiality classification and that can be published to the public.

Research data is located between the bottom and the middle level. In some cases, measurement data is highly confidential and cannot be shared among the members of the project consortium directly. A concrete example for this kind of data is personal data according to the GDPR. In other cases, the confidentiality is not as high, so that the data can be shared with the other project members. Concrete examples for this kind of data are measurement data from performance tests, energy monitoring or point clouds. Therefore, the category of research data is situated between the bottom and middle level of confidentiality. Data from the category analyzed research data is located between the middle (shareable among project consortium) and highest (shareable with the public) level. Some of the research results will be published which require anonymization of the respective data. Data from the third category – reports and communication – is at the highest level enabling the publication of the data to the open public. The final category containing the internal project data is located between the lowest and middle level. Most project data can be shared among the members of the project consortium. However, sensitive data and information like detailed financial information will not be shared among all

consortium members. Next to the project consortium, the projects that are funded via financial support for third parties (FSTP) will also generate data. This data is categorized equally to the data of the project consortium and will only be shared or made available according to the levels of confidentiality.

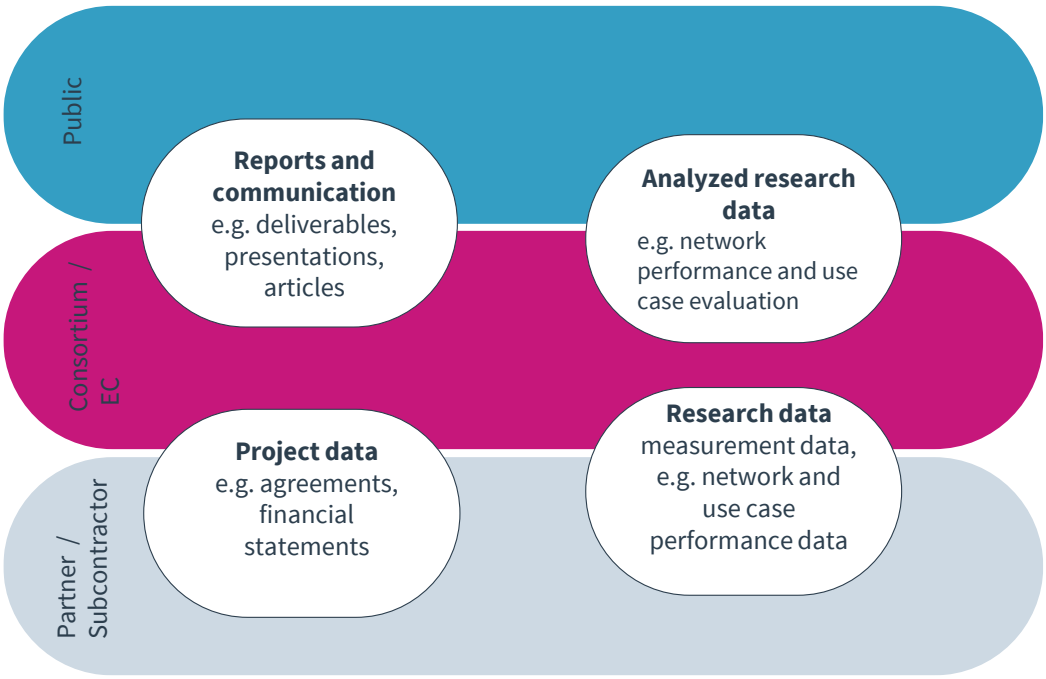

Figure 1: Categories of Data based on [5]

2.2 Research data

The TARGET-X consortium will generate experimental data from the use cases and consider reusing existing data to enhance the research outcomes. By leveraging relevant pre-existing data, the project aims to enhance the depth of its research outcomes. The project itself is expected to generate and utilize big datasets, amounting to approximately 15 TB of data in total. These datasets will be crucial in conducting in-depth analysis and validating the experiments outlined in the project's work plan. The origin/provenance of the data can be internal to the project (generated by project activities) or external (reused from existing sources). Specific data sources or providers will be determined during the project.

Projects within the Stream D of the 6G Smart Networks and Services (SNS), such as TARGET-X, have a focus on measurement and validation actions, and thus most of the generated research data is expected to be measurement data sets of various kinds and origin.

Exemplary research data sets that may be generated or reused during the project include:

- Data from KPI and KVI analysis with the Methodological Assessment Framework from WP1 including product data, process data, failure data, facility data, sustainable data, safety data and security data
- Localization data and communication data for robotics as well as for manufacturing will be generated in WP2 Manufacturing

- Energy monitoring data and data from tests in WP3 Energy
- Localization data for automotive, network KPIs as well as signal level and coverage data from WP4 Automotive
- Location data, machine and network status information, point cloud scans, video recordings, lidar scans, thermal maps or infrared scans, radar scans, building element passports, BIM models, planning and logistics data and performance data in WP5 Construction
- Performance measurement data, measurement execution protocols, coverage measurements, AAS submodels from WP6 Technology evolution

More details regarding the purpose, origin, format, size, reuse indication and availability of each dataset mentioned above can be found in Annex 1 of this document.

At this stage in the project execution, we foresee that potentially also research data other than (raw) data sets will be generated. For TARGET-X, these categories of additional research data can be identified:

1. Connectivity solutions: in the execution phase of the project, new solution to connect domains that currently not connected will need to be developed to meet the ambitions of the project. These solutions could be definitions of workflows on how to interconnect the entities, software modules or other bits of software.
2. Anonymization scripts/workflows: data sets from network captures with tools, such as Wireshark, are likely to contain information that should not be publicly available. Scripts and/or workflows will need to be created to remove this information from raw data sets.
3. Conversion scripts: analysis and correlation of data from different data sets often requires the conversion of data points within these sets, to unify the data points across multiple data sets of different origin.
4. Simulation models: for activities such as radio propagation calculations, realistic models of environments are needed. It is expected that these models will have to be created.

Other research output generated by TARGET-X includes meeting minutes from the different TARGET-X bodies and work packages, as well as administrative data for the project execution. Further administrative data, such as applications and surveys, is generated through the execution of the third-party projects. This data is part of the project data category and will not be shared publicly. TARGET-X also produces information material and other data related to communication and dissemination, such as web analytics and statistics from social media. If applicable, they will be shared through deliverables.

2.3 Ethical Aspects

Ethics and legal issues that impact the sharing of datasets must be considered in a DMP, since data management is closely intertwined with ethical considerations and legal frameworks, especially when dealing with sensitive or personal data. Consequently, ethical data management requires obtaining informed consent from individuals whose data is being collected, processed, or shared. Therefore, in TARGET-X, participants and companies will be fully informed about the purpose, risks, and potential benefits of data sharing, and that they have the choice to give or withhold consent.

The sharing of datasets in TARGET-X should also protect the privacy and confidentiality of individuals. It is crucial to anonymize or pseudonymize personal data to minimize the risk of re-

identification. This ensures that individuals or companies cannot be directly or indirectly identified from the shared data.

Additionally, in TARGET-X, personal data is only used for organizational purposes, e.g., for the realization of the open calls for The Financial Support for Third Parties (FSTP). Personal data will not be shared with unauthorized persons and only in compliance with GDPR requirements. Therefore, appropriate security measures will be implemented to protect the integrity and confidentiality of shared data. Including informed consent in a questionnaire is currently out of the scope of the project. If personal data becomes relevant for other purposes, including informed consent into the questionnaire will be considered.

Furthermore, an ethics review may be required before initiating data collection and sharing. This review should ensure that the research complies with ethical principles and legal requirements. The ethics chapter in the Description of the Action (DoA) outlines the ethical considerations specific to the project and provides guidance on how to address them.

3 Making the data FAIR

FAIR data is findable, accessible, interoperable, and reusable. These are the four foundational principles for data management, and they are a tool to maximize the value created by the data.

The FAIR Guiding principles from [3] are the following:

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
- A1.1 the protocol is open, free, and universally implementable
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
- R1.1. (meta)data are released with a clear and accessible data usage license
- R1.2. (meta)data are associated with detailed provenance
- R1.3. (meta)data meet domain-relevant community standards

The FAIR principles will be referenced during this chapter.

3.1 Findability

The findability will be addressed in TARGET-X using identifiers, the provision of metadata and relevant keywords.

3.1.1 Identifiers

Identifying data with a persistent identifier is crucial for findability. If data is published, it will be associated with a Digital Object Identifier (DOI) that also allows for referencing the used data set in a research paper.

Next to the data, also the meta data of the data will be assigned persistent identifiers. Those identifiers include:

- The TARGET-X Project DOI https://doi.org/10.3030/101096614 that links to the project description of the European CORDIS page
- Open Researcher and Contributor IDs (ORCIDs) [6] for identifying the researchers involved in the data creation. The use of ORCIDs for identifying the involved researchers is encouraged. ORCIDs will be associated to publications and published datasets

The use of persistent identifiers for the data, the project and the researchers contribute to the principle F1.

Documents internal to the consortium have no persistent identifier but are referenced with a joint naming convention and a filesystem that is similar for each work package of the TARGET-X project.

3.1.2 Meta data

Metadata has a significant impact on the findability of datasets. To facilitate automatic discovery of the public datasets that will be created during the TARGET-X project, machine-readable metadata according to open-standard vocabularies will be provided along with those datasets. Exemplary open-standard vocabularies for describing digital and physical resources that fulfill the Horizon Europe 2020 requirements are the Dublin Core terms maintained by the Dublin Core™ Metadata Initiative or the RADAR generic metadata schema [7]. These metadata schemas support both machine processing as well as human recognition. Thus, they satisfy the FAIR principles regarding findability while at the same time laying the foundation for the FAIR principles regarding reuse. To promote usability and application of the suggested schema within the TARGET-X consortium, metadata generators such as provided by Coscine [8] will be used. Furthermore, all published metadata will include a DOI as persistent identifier as described in section 3.1.1. Providing rich metadata contributes to the FAIR principles F2, F3 and R1.

Metadata is an important element to allow the reuse or verification of provided data sets. Within this project the metadata will be licenced under the CC0 if not specified differently (FAIR principle R1.1). The metadata can be either stored embedded in the data file, for example in the HDF5 format or alongside. In both cases usually the data and metadata need to be seen as one and cannot be used without the other. Therefore, most of the time the metadata will expire together with the data. In some exceptional cases it could make sense to keep metadata accessible after the lifetime of the use data. How to handle these cases will be decided on a case-by-case basis by the data provider. A second key element concerning the reusability of any type of data is the data encoding and the description of the software needed to allow for the decoding of specific data sets. As far as possible this project will rely on open source or publicly known formats such as CSV, HDF5, DOCX and others. In case of a specific format that is not commonly known further instructions on how to decode the data will be provided either within the metadata of the data set or this document.

For published data sets, Zenodo stores metadata (and data) for the lifetime of the repository and provides the Open Archives Initiative Protocol for Metadata Harvesting API (OAI-PMH) this allows to harvest the metadata of the repository. This contributes to FAIR principles A1 and A2.

3.1.3 Keywords

Keywords will be provided in the metadata to optimize the discovery and potential re-use of the published data. Relevant keywords are also associated with the TARGET-X project description on the European CORDIS website [9].

Figure 2: Keywords on the TARGET-X CORDIS website [9]

The keywords named there, are a base for the keywords for the data sets. The data set's meta data will include a relevant subset of the TARGET-X keywords and further keywords related more closely to the purpose and nature of the respective data set.

3.2 Data storage

An important component of the FAIR principle is to ensure that acquired data is available to project partners but also to third parties. Based on the type of data and the scientific field in which the data was collected, different methods of sharing are commonly used. This section therefore first discusses the means of storing documents and other documents and then will discuss the storage of other larger datasets. It is worth mentioning that even though the project aims at providing as much data as possible for the public it still must be decided for each case individually if datasets can be published. This applies to all kinds of datasets within the project. The reasons for not publishing a dataset could vary largely. Some examples are that the data would disclose protected intellectual property, another one would be that the data reveals information about critical infrastructure e.g., the power grid. Any set of data collected and intended to be made available on a public platform, must have a broadcasting authorization by the operator of the site of origin.

3.2.1 Repository

This section focuses on the storage of measurement data as for that larger data sets are expected. This could include any type of data that is acquired within a measurement campaign. This type of data will, in most cases, first be stored on premise and depending on the data a subset or revised set of the original data is published. The goal is to allow for minimal effort for the potential data consumer, therefore the augmentation of the data is a very important step. For publication different

platforms are available. In most cases such repositories are not peer reviewed. In the following some commonly used platforms are described. Depending on the use case a different, better suited, platform could be chosen. An extensive list of other platforms can also be found at [10] or [11].

GITHUB/GITLAB

GitHub [12] and GitLab [13] are two examples for repositories based on the GIT versioning technology. This kind of repository is mostly used for collaborative work on source code including an automatic versioning of the data and the possibility for running automated routines as soon as a new version is available. Such routines could be automated code testing, a build process or code analysis. The above-mentioned repository providers already provide a space for a large community of open-source projects. Both allow also for private projects which allows for development and then publication of the final version. It is worth mentioning that GitLab can also be self-hosted and in fact is adopted by most of the German universities and research organizations. Therefore, these local hosted platforms also allow for collaborative work and publication of the repository by, at the same time, ensuring that the actual data is stored within the EU.

ZENODO

Zenodo [14] is a platform for long term storage of data. The service is provided by the research facility CERN and financed by the EU. For data sets published via Zenodo a DOI can be generated. The DOIs can also be generated before publication, so data sets can be referenced in a peer-reviewed article and jointly published with it. The platform can handle single datasets with up to 50 GB size. At this point of the project, the data size limit seems to be sufficient. If more storage per dataset is needed, CERN encourages the user to contact them. The amount of space is not limited [18]. The platform intends to help research project to share data all over the world. To do so the platform also helps by defining and storing some additional metadata provided by the uploader. It is possible to grant access to the data only to a specific group of users or the public. The platform also gives the user the possibility to set an embargo period for the data.

3.2.2 Open data and restrictions

The guiding principle is to make data openly available as much as possible within the project's activities, keeping within the borders agreed in the consortium and described in the grant agreement. The datasets, generated in TARGET-X and marked for publishing, will be uploaded and stored in research data repositories. A preference for Zenodo is given, but other repositories can be used as well. If data is secret or confidential, Zenodo is not suitable for storing the data [15] . In this case, the European Commission's principle 'as open as possible, as closed as necessary' will come to play [16]. For most of the datasets, open access shall be used. This access method provides an unrestricted and barrier free access to the dataset. The use of embargoed access is recommended in case of :

- Publications: the embargo end date on the research data shall be selected to match the date of appearance of the associated text publication
- Intellectual Property Rights: the chosen embargo end date shall define a period that is equal or longer than the agreed protection period

In case there are doubts on the access type to be selected, the upload of the dataset shall be postponed until the suitable access type is fully clarified. The use of anonymizer scripts or tools, such as “amnesia” [17] prior to uploading any dataset is recommended.

3.2.3 Data security

In general, a suitable data storage solution is needed as soon as any kind data is handled. This also includes backup and restore strategies for important data sets. The period for which data is backup needs to be determined on based on the data sets. All partners are responsible to enforce such strategies for the relevant datasets. This could either be done by providing an inhouse infrastructure to allow for backup and restore or could be outsourced by utilizing a storage provider as discussed in section 3.2.1. Not all collected data sets are going to be published or even evaluated and stored. Often during prototyping and development data is acquired to allow for a fast development cycle. Therefore, for this type of data there is no need to invoke any backup strategy. Which data needs to be backed up or even stored is decided by the data provider.

3.2.4 Storage of internal documents

Storage and management of internal documents is implemented using Fraunhofer's MS SharePoint instance. This SharePoint acts as a central point of communication and file distribution among the project consortium and enables joint and simultaneous editing of data and documents. In this way, project members can collaborate more efficiently e.g., during the writing of phase of project deliverables. The SharePoint provides versioning and is managed by Fraunhofer IPT. Additionally, Fraunhofer IPT also maintains a network drive on its own servers that acts as a back-up to SharePoint. All relevant internal documents (deliverables, presentations, important meeting notes, etc.) are stored on this network drive to ensure that Fraunhofer IPT has full control over and access to all relevant data and documents. The data stored on the network drive is archived with snapshots over a period of thirty days on an hourly basis, so that project members also have access to older version of data and documents. After the project’s completion, the entire project directory is archived for long-term preservation using a magnetic tape data storage system (see also section 3.2.5).

3.2.5 Longterm preservation

Project directories containing all relevant project data are archived and stored by Fraunhofer IPT for a period of ten years. In case a preservation of certain data beyond the usual ten years is required (for instance because project members consider the long-term preservation of certain datasets necessary), the project management team will discuss the matter and take actions to preserve the data for a longer period. Fraunhofer IPT employs magnetic tape data storages for long-term data preservation. In this way, a long-term preservation of relevant data from the TARGETX project is ensured. Since Fraunhofer IPT stores all its project data in this way, the technical infrastructure already exists, so that there are no significant additional costs. If the preservation period is to be extended, no additional measures must be taken.

Long term preservation of published data sets and their meta data will be ensured by the chosen repositories. The targeted repository is Zenodo. Zenodo is a trusted repository and will store the data as long as it exists and migrate the data to suitable repositories in case it must close [18].

3.3 Making data interoperable

All TARGET-X partners will ensure the interoperability of published datasets and the respective metadata by using common file formats, standards, vocabularies and methodologies. A detailed collection of the used as well as expected file formats is given in section 2.1 and Annex 1. Generally, documents will be provided in standard formats (e.g., DOCX, XLSX, PDF) to ensure accessibility and usability by all partners. In contrast to this, for technical data dominant exchange formats such as JSON, XML, CSV etc. will be utilized. For data processing and analysis open-source software is going to be used whenever possible. Metadata will be generated according to an established schema as described in section 3.1.2. If the use of non-standard vocabularies or ontologies proves unavoidable, or if custom vocabularies and ontologies will be developed during the project, they will be made available along with mappings to standard vocabularies and ontologies.

3.4 Increase data re-use

The fourth pillar of the FAIR approach is the assurance of data re-usability. Subsequent projects in the Smart Networks and Services Joint Undertaking (SNS JU), or other research programs, will benefit from the availability of accessible datasets for within the scope of their activities.

The successful re-usage of research data relies on the quality of the datasets and even more on the quality of the associated metadata and descriptions of the shared datasets. TARGET-X will use a peer review principle to perform quality assurance tasks for uploaded datasets.

3.4.1 Data description

The datasets will have properly defined metadata associated to them in a format supporting easy machine-based analysis. In addition, a description of the dataset, detailing what data was collected, how the data was acquired, why the data was acquired, who acquired the data shall be stored with the dataset as a text file, written in human-readable format.

The details on metadata are described in section 3.1.2.

3.4.2 Data sharing

Sharing and using shared data can only be done, provided that the datasets have been made available under the proper licensing terms. Datasets, uploaded to repositories such as Zenodo, will remain available after the termination of TARGET-X. For datasets that are shared via other means, no provisions are taken to assure availability after the project has ended.

The use of datasets generated outside of TARGET-X activities is considered beneficial to the results generated by TARGET-X. The supplemental datasets would increase the available datapoints for analysis and performing analysis on datasets collected in different settings allow to draw more generic conclusions. The datasets that are used for such activities shall be documented, along with their provenance.

4 Conclusions

The TARGET-X DMP used usual practices in data management to identify the best way to store, maintain, and disseminate scientific information. In this DMP, we showed the data management practices and procedures that will be used in the project. In addition, we showed how the data will be stored and presented with examples for metadata, data formats, and data types. The deliverable also presented the different options that the project is considering providing FAIR and free of charge access to the general scientific information generated by the partners with an explanation of the assigned resources.

This document will be a living document and will be updated based on the changes that may occur in the type and amount of generated scientific information.

5 References

[1] https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan_he_en.docx
[2] https://argos.openaire.eu/splash/
[3] Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, I. Jsbrand Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie et al. (2016): The FAIR Guiding Principles for scientific data management and stewardship. In: *Sci Data* 3 (1), S. 160018. DOI: 10.1038/sdata.2016.18.
[4] https://nfdi4ing.de/
[5] IIT Deliverable. D1.2 Data Management Plan. Online available at https://services.phaidra.univie.ac.at/api/object/o:1139680/download, last accessed 22.05.2023.
[6] https://orcid.org/
[7] https://radar.products.fiz-karlsruhe.de/en/radarfeatures/radar-metadatenschema
[8] https://coscine.de/
[9] https://cordis.europa.eu/project/id/101096614
[10] https://data-collections.nfdi4ing.de/
[11] https://www.re3data.org
[12] https://github.com/
[13] https://gitlab.com
[14] https://zenodo.org/
[15] https://about.zenodo.org/infrastructure/
[16] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm
[17] Amnesia Anonymization Tool website, visited May 2023 https://amnesia.openaire.eu/index.html
[18] Zenodo - Research. Shared. Frequently Asked Questions (2023). Online available at https://help.zenodo.org/, last accessed 23.06.2023.
[19] https://developers.zenodo.org/#oai-pmh

Annex 1 Exemplary data sets

The entries in the table below can be understood as examples of the datasets that may be generated or reused during the project. While efforts have been made to provide a comprehensive overview, it is essential to note that the actual datasets and their characteristics may vary. This table aims to illustrate the variety of data that could be involved and highlight their potential usefulness beyond the project.

Table 1: Exemplary TARGET-X data sets

N°	DATASET	WP	PURPOSE	ORIGIN	FORMAT	SIZE	REUSED	USEFUL FOR	AVAILABLE AFTER /OUTSIDE THE PROJECT
01	**Product data (KPI)**	1	Description of quantity, quality, and type of products, including construction materials.	-	Text	-	No	Construction companies, architects, building material suppliers	After project completion
02	**Process data (KPI)**	1	Characterisation of process/application quality and type.	-	Text	-	No	Project team, researchers	During and after project
03	**Failure data (KPI)**	1	Characterisation of unplanned and undesirable events.	-	Text	-	No	Project team, researchers	During and after project
04	**Facility data (KPI)**	1	Describes trial sites (production sites, construction sites, etc.) and resource consumption (material and energy).	-	Text	-	No	Project team, stakeholders	During and after project
05	**Sustainability data (KVI)**	1	Data on resource consumption (material and energy) for all verticals.	-	Text	-	No	Construction companies, researchers	During and after project
06	**Safety data (KVI)**	1	Data describing safety aspects, especially in manufacturing, construction, and autonomous	-	Text	-	No	Construction companies, safety experts	During and after project

			driving verticals.						
07	**Security data (KVI)**	1	Data describing IT-security aspects.	-	Text	-	No	IT professionals, security experts	During and after project
08	**New value chain data (KVI)**	1	Integration of data from different product life cycle phases for all verticals.	-	Text	-	No	Project team, researchers	During and after project
09	**Localisation data for robotics**	2	Ground truth localisation data for assembly logistics task.	-	Text, .log	-	No	Robotic system developers, logistics experts	After project completion
10	**Communication data for robotics**	2	Communication statistics during robotic bin-picking use case or whole-body movements of mobile manipulators.	-	Text, .log	-	No	Robotic system developers, communication experts	After project completion
11	**Localisation data for manufacturing**	2	Localisation data with time stamps for workpieces and tools with process and layout descriptions.	-	Text	-	No	Manufacturing, logistics companies	After project completion
12	**Communication data for manufacturing**	2	Communication statistics (e.g., latency, bandwidth) during BLISK manufacturing with inline quality insurance.	-	Text	-	No	Manufacturing, quality assurance experts	After project completion
13	**Energy monitoring data**	3	Data on energy, power, voltage, and current in ARGOS format.	-	Text	-	No	Energy analysts, researchers, utilities	During and after project
14	**Data from tests**	3	Data from tests as described in the deliverable description.	-	Text	-	No	Project team, researchers	During and after project

15	**Localisation data for automotive**	4	Location of vehicules with time stamps	GPS	Text	-	No	Road operators, network engineers, automotive application developers	During and after project
16	**Network KPIs**	4	Edge/cloud downlink and uplink throughput, edge/cloud latency.	Log files	Text	-	No	Telecommunication companies, Network engineers, system administrators, automotive application developers	During and after project
17	**Signal level and coverage data**	4	-	sensors	Text	-	No	Telecommunication companies, network analysts	During and after project
18	**Localisation data**	5	Data of machines and building elements with GPS and RTK refinement.	-	Text	-	No	Construction companies, building owners	After project completion
19	**Kinematic data**	5	Kinematic data of construction machines for motion and path planning.	-	Text	-	No	Construction companies, robotics researchers	After project completion
20	**Machine status**	5	Such as battery status.	-	Text	-	No	Equipment manufacturers, maintenance teams	After project completion
21	**Server up/downtimes**	5	Data on the server up and downtime.	-	Text	-	No	IT professionals, system administrators	After project completion
22	**Point cloud scans**	5	The scans of construction sites and building elements in.ascii or .lxs format.	-	Text	-	No	Architects, engineers, construction companies	After project completion
23	**Video recordings**	5	Machine navigation, object detection, and surveillance cameras.	-	Video	-	No	Robotics researchers, security experts	After project completion

24	**LiDAR scans**	5	Point clouds from LiDAR scanners.	-	Point cloud	-	No	Surveyors, geospatial analysts	After project completion
25	**Thermal maps or infrared scans**	5	-	-	Image	-	No	Energy analysts, architects, building owners	After project completion
26	**Radar scans**	5	Point clouds from Radar scanners.	-	Radar data	-	No	Radar technology experts, researchers	After project completion
27	**Building element passports**	5	Data including production date, material composition, and delivery status of building elements.	-	Text	-	No	Construction companies, building owners	After project completion
28	**BiM-model**	5	BiM-model including architectural design and structural design in .ifc files.	-	.ifc files	-	No	Architects, engineers, construction companies	After project completion
29	**Literature**	5	-	-	Text	-	No	Researchers, scholars	During and after project
30	**Planning and logistics data**	5	Including orders, deliveries, shipment info, calendar entries, etc.	-	Text	-	No	Project managers, logistics companies	During and after project
31	**Energy monitoring data**	5	Energy monitoring data in .ascii format (~500MB).	-	Text	~500 MB	No	Energy analysts, researchers, utilities	During and after project
32	**Performance tests**	5	Performance tests of the 5G campus network, including QoS and communication statistics.	-	Text	-	No	Network engineers, system administrators	During and after project
33	**Scrap rate**	5	Scrap rate of building processes.	-	Text	-	No	Construction companies, quality control experts	During and after project

34	**Reusability rate**	5	Reusability rate of materials/building elements.	-	Text	-	No	Construction companies, sustainability experts	During and after project
35	**Performance measurement data**	6	Performance measurement data.	-	CSV, PCAP	-	No	Researchers, analysts	During and after project
36	**Measurement execution protocols**	6	Measurement execution protocols.	-	Text	-	No	Researchers, analysts	During and after project
37	**Coverage measurements**	6	Coverage measurements and plots.	-	Text,PNGS	-	No	Network engineers, researchers	During and after project
38	**AAS sub-models**	6	-	-	Text	-	No	Researchers, system developers	During and after project