transformer

Data Management Plan

Deliverable No.:	D1.1
Project Acronym:	TRANSFORMER
Full Title: Data Management Plan	
Designing long-term systemic transformation frameworks for regions. Accelerating the shift towards climate neutrality	
Grant Agreement No.:	101069934
Work package/ Measure No.:	WP1
Work package/ Measure Title: Project Coordination & Management	
Responsible Author(s): Morgane Juliat (RC)	
Responsible Co-Author(s): Sounak Bandyopadhyay (RC)	
Date:	15.12.2022
Status:	Submitted
Dissemination level:	Public

Abstract

The Data Management Plan to the granting authority in accordance with the Grant Agreement, which summarises how the project will ensure open access to research data and will secure usage of personal data according to EU and national laws.

Project Partners

Organisation	Country	Abbreviation
RUHR-UNIVERSITAET BOCHUM	DE	RUB
RUPPRECHT CONSULT-FORSCHUNG & BERATUNG GMBH	DE	RC
BUSINESS METROPOLE RUHR GMBH	DE	BMR
REGIONE EMILIA ROMAGNA	IT	RER
FONDAZIONE ISTITUTO SUI TRASPORTI E LA LOGISTICA	IT	ITL
FIT CONSULTING SRL	IT	FIT
Dolnoslaski Fundusz Rozwoju sp. z o.o.	PL	DFR
UNIWERSYTET WARSZAWSKI	PL	Uni Warsaw
Fundacja Dumni z Lubina	PL	Dumni z Lubina
ANKO DYTIKIS MAKEDONIAS A.E. - ANAPTYXIAKOS ORGANISMOS TOPIKIS AFTODIIKISIS	GR	ANKO
ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS	GR	CERTH
TWENTY COMMUNICATIONS SRO	SK	TWE
EUROPEAN NETWORK OF LIVING LABS IVZW	BE	ENoLL

Document History

Date	Person	Action	Status	Diss. Level
02.12.22	Morgane Juliat (RC)	Submission of the document to reviewers	Draft	WPL
09.12.22	Georgia Ayfantopoulou (CERTH)	Review	Draft	WPL
09.12.22	Maria Konstantinidou (CERTH)	Review	Draft	WPL
11.12.22	Thomas Meister (RUB)	Review	Draft	WPL
15.12.22	Morgane Juliat (RC)	Final Review	Final	WPL
15.12.22	Thomas Meister (RUB)	Approval	Approved	PC
		Submitted		PO

Status: Draft, Final, Approved, and Submitted (to European Commission).

Dissemination Level: WPL = Work Package Leader, PM = Project Manager, PC = Project Coordinator, PO = Project Officer

Funded by
the European Union

Legal disclaimer

The sole responsibility for the content of this deliverable lies with the authors. It does not necessarily reflect the opinion of the European Union. The European Commission is not responsible for any use that may be made of the information contained therein. All images are provided by the respective partners (unless otherwise noted) and are approved for reproduction in this publication.

Funded by
the European Union

Table of Contents

Table of Tables

Deliverable executive summary

This document describes the Data Management Plan (DMP) for the TRANSFORMER project and is produced as part of WP1. The DMP provides an analysis of the main elements of the data management policy that will be used throughout the project by the project partners, concerning all the datasets that will be generated, harvested and/or used by the project.

In more detail this document covers the following:
1. TRANSFORMER data identification and collection approach
2. TRANSFORMER overall dataset structure, including an overview of identified data sources and datasets
3. The TRANSFORMER overall data management plan
4. Dataset with personal data in TRANSFORMER
5. Ethical aspects related to data management in the TRANSFORMER project.

As data management is an ongoing process throughout the TRANSFORMER project and data management in the project is taking place in a dynamic environment, this document is seen as a living document, and as such will follow and reflect the evolution in the form of dataset updates and/or respective changes in the Consortium policies.

Chapter 1: Introduction

Reaching the goal of net-zero emissions by 2050 requires immediate actions going beyond the level of fostering innovation and digitalisation in societal niches. In fact, it calls for an innovation path which sets out to 'design' carbon-neutral societal systems, and focuses on investments in zero-carbon solutions as indicated in the "Final Report of the High-Level Panel on Decarbonisation 2018".

TRANSFORMER takes up this challenge by applying the Transition Super-Lab (TSL) approach as "very-large territory initiatives of real-life management of the transition from typical fossil-fuel-based local economies to zero carbon ones[1]". Super-Labs combine the evidence-based success of fast and effective decarbonisation processes with real-life, and large-scale development and testing of portfolios of low-carbon, cost-effective and proven technological (and non-technological) solutions.

This document describes the Data Management Plan (DMP) for the TRANSFORMER project. This document aims to provide an overview of the main elements of the data management policy applied in this project, concerning all the datasets that will be generated, harvested and/or used by the project. This

---

[1] Final Report of the High-Level Panel of the European Decarbonisation Pathways Initiative (2018), p. 165. Available at: https://ec.europa.eu/info/publications/final-report-high-level-panel-european-decarbonisation-pathways-initiative_en

deliverable provides the first version of the DMP and should be considered a living document and will be updated during the project's lifetime.

TRANSFORMER will publish its results on the Open Research Europe platform and follow thoroughly all the required actions to be aligned with the Open Science practices as they are defined in Horizon Europe guidelines. TRANSFORMER will take an open science approach to develop its methodology using open standards, protocols and, where applicable, open-source software. The project manager will be responsible to set up an adequate internal data management structure to ensure the protection of internal and confidential data. Research data collected and generated by the project will be opened and offered to the Open Research Data Pilot, in which TRANSFORMER declares its intention to participate.

The consortium is committed to making knowledge, information and data produced by this project exploitable and accessible as much as possible to third parties. Their disclosure will be made considering the confidentiality of some data, necessary to protect the competitiveness of the General Data Protection Regulation (GDPR). The protection of intellectual property will follow the principles of the European Union and will be addressed in the consortium agreement. The majority of the project's research output will be published following open-access requirements.

Chapter 2: Data Summary

Towards this direction, TRANSFORMER will provide methodologies and tools which will be produced through data collection. The innovative nature of TRANSFORMER, researching the application of transition super-labs with an explicit focus on simultaneously transforming more than one societal system, creates complexities in the identification of all the data that will be used at this stage of the project. It is too early in the project to define in detail all required datasets for the implementation of all TSLs and more datasets will be added during the project. A preliminary estimation of the required datasets for each WPs can be found in the Annex.

At this stage of the project, the following type of data has been identified:

- Database of TSL predecessors – will contribute to the development of the TSL concept and contributes to a common understanding of the methodological approach and conceptual grounding of evaluation and impact assessment within the TRANSFORMER project.
- Database of regions across Europe with a potential need for the TSL approach
- Summary of expert interviews and document analysis for the feasibility studies (mainly qualitative data, supplemented with quantitative data about the regions)
- Database of TSLs internal and external stakeholders (Personal data will be anonymized or only published with the explicit approval of the person (we collect names, the professional position of the persons etc.).)
- Metadata for the Emilia-Romagna (provided by the DCAT-AP_IT model with additional metadata, according to each own necessity albeit in full compliance with the rules as defined in

the DCAT AP_IT specification). The scope of the data collection/generation for the TSL of Emilia-Romagna is to understand the state-of-art within the existing mobility and energy plans and contribute to the improvement and harmonisation of different sectors within society serving environmental, economic and social needs and objectives.

- Type of data includes the open format, MTS, data from questionnaires, data from cities
- Database on mobility and energy for Western Macedonia (in multiple data formats). Dataset will be accompanied by detailed documentation of its contents. Data collection is necessary for the implementation of the 3 use cases in Western Macedonia.

The data listed above may come from different sources, which have been identified below:

- Internet (desk research)
- Focus groups
- Expert interviews (using anonymization techniques)
- Document analysis (e.g. SUMPs, regional plans and studies such as Electric Vehicle Charging Plan, Regional mobility plan, Sustainable Energy Action Plan etc)
- Data gathering workshop
- Survey for stakeholders
- Surveys that consider the needs of the citizens
- Mobility and energy ecosystem (businesses, transport modes and private providers, land use, etc), transport operators and other stakeholders related to each use case

Table 1: Milestones and deliverables in relation to data gathering or generation

WP	Milestone/Deliverable	Due	Data Type	Responsible Partners	Data access level
2	D2.1 Summary of data collection on TSL predecessors	31.10.22	Report in .pdf	RUB	Public
2	D2.2 Quantitative mapping research report	28.02.2023	Report in .pdf	RUB	Public
2	D2.3 Regional SWOT analyses as feasibility studies to be used as evidence-based in decision-making for roadmap construction	31.08.2023	Report in .pdf	RUB	Public
2	D2.4 Academic working paper(s) for conceptual framing of Transition Super Labs	31.08.2024	Report in .pdf	RUB	Public

Funded by
the European Union

WP	Milestone/Deliverable	Due	Data Type	Responsible Partners	Data access level
3	D3.1 Recommendations for Transition Super-Lab coalitions building, empowering of vulnerable and marginalised groups, and vision process	30.04.2023	Report in .pdf	ENoLL	Public
3	D3.2 Definition of Transition Super-Lab use cases	31.05.2023	Report in .pdf	BMR	Public
3	D3.3 Transition Super-Lab Action Plan (APs per TSL as Annex)	30.04.2024	Report in .pdf	BMR	Public
3	D3.4 Transition Super-Labs' Lessons Learned	31.08.2024	Report in .pdf	CERTH	Public
4	D4.1 Super-Lab Roadmap - version 1	31.08.2023	Report in .pdf	RC	Public
4	D4.2 Super-Lab Roadmap – version 2	31.08.2024	Report in .pdf	RC	Public
4	D4.4 Knowledge Hub	31.10.2023	Report in .pdf	CERTH	Public
5	D5.1 Framework for Super-Labs Assessment - version 1	31.05.2023	Report in .pdf	CERTH	Public
5	D5.2 Framework for Super-Labs Assessment - version 2	30.06.2024	Report in .pdf	CERTH	Public
5	D5.3 Best practices and recommendations for Super-Labs operation towards the region transition	30.06.2024	Report in .pdf	CERTH	Public
5	D5.4 Guidelines for Super-Labs tools	30.06.2024	Report in .pdf	FIT	Public
6	D6.1 Communication and Dissemination Strategy and Plan - version 1	30.09.2022	Report in .pdf	TWE	Public
6	D6.2 Communication and Dissemination Strategy and Plan - version 2	30.11.2022	Report in .pdf	TWE	Public

WP	Milestone/Deliverable	Due	Data Type	Responsible Partners	Data access level
6	D6.3 TRANSFORMER Hub	28.02.2023	Report in .pdf	TWE	Public
6	D6.4 TRANSFORMER capacity building	31.08.2024	Report in .pdf	ENoLL	Public
6	D6.5 TRANSFORMER Exploitation Plan	31.08.2024	Report in .pdf	TWE	Public

Contract for data collection

In terms of specific contracts related to the process of collecting data, there is only one case, for Western Macedonia, where a subcontract will be signed between CERTH and a subcontractor for the mobilisation of local stakeholders to be involved, in order to gather experience & contribute towards designing and implementing the pilot actions & the development of the TSL in Western Macedonia. This contract will be developed following a public procurement process.

Chapter 3: Fair Data

A Data Management Plan (DMP) is developed in WP1 ensuring that all research outputs and data in TRANSFORMER are Findable, Accessible, Interoperable and Reusable (FAIR). As indicated, FAIR principles would be achieved through Persistent data Identifiers (PIDs) and Digital Object Identifiers (DOIs)); deposit in trusted open data repositories such as Zenodo; open, free-to-use tools and documents, ensuring one of the core targets of TRANSFORMER so that not only TRANSFORMER TSL regions but also more regions in the future have access to tools to achieve their long-term systemic transformation towards climate neutrality. The next paragraph describes in detail the FAIR principles.

Making Data findable, including provisions for metadata

Data are **Findable** when they are described by sufficiently rich metadata and registered or indexed in a searchable resource that is known and accessible to potential users. Additionally, a unique and persistent identifier should be assigned so that the data can be unequivocally referenced and cited in research communications. The identifier enables persistent linkages to be established between the data, metadata, and other related materials in order to assist data discovery and reuse. Related materials may include the code or models necessary to use the data, research literature that provides further insights into the creation and interpretation of the data and other related information.

To ensure the findability of data, partners in the TRANSFORMER project will use clear version numbers. Some of the data will follow a specific naming convention, such as the naming convention of national

profiles (i.e. DCAT-AP_IT). In some cases, search keywords will be provided in the metadata to optimize the possibility for discovery and then potential re-use.

Making data accessible

Accessible data objects can be obtained by humans and machines upon appropriate authorisation and through a well-defined and universally implementable protocol. In other words, anyone with a computer and an internet connection should be able to access at least the metadata. It is important to emphasise that Accessible in FAIR does not mean Open without constraint. Accessibility means that the human or machine is provided – through metadata – with the precise conditions by which the data are accessible and that the mechanisms and technical protocols for data access are implemented so that the data and/or metadata can be accessed and used at scale, by machines, across the web.

TRANSFORMER will address

1. early and open sharing of research through preregistrations of the research plans in advance of the study implementation, as well as registered reports, will be submitted to Open Access Repositories;
2. Research output management and measure to ensure reproducibility of research output – TRANSFORMER will be careful to ensure the reproducibility of the outputs by covering the three main research processes that reproducibility is based on (reproduction, replication and re-use);
3. Open access to research outputs and participation to open peer-review: relevant TRANSFORMER publications will be published in Open Access Journals which will be initially checked from the SHERPA/RoMEO platform and the Directory of Open Access Journals (DOAJ)in order to confirm the open access and copyright policies of specific international journals, and **Open Access Repositories** (mainly Green Open Access journals and Gold Open Access journals or e.g. Zenodo, arXiv, etc.) which will be identified through platforms as Risk Observation, Assessment and Remediation (ROAR), OpenDOAR, OpenAiRE and OAD;
4. Accessible results and deliverables will be published on the TRANSFORMER Hub (project's website) will serve as the central communication channel for the project and disseminated to the relevant target groups through various communication measures (newsletter, social media etc.);
5. Involvement of relevant knowledge actors including civil society, citizens and relevant target groups in the co-creation of R&I agendas and contents.

The term open access (OA) concerns the free online provision of re-useable (scientific) information to other users. The scope of publicly funded research and innovation projects and incentives is to contribute to the improvement of different sectors within society serving environmental, economic and social needs and objectives. The benefits of open access to scientific outcomes are outlined in the "Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020" and are described briefly below:

- Accessible existing scientific publications and results form the basis for efficient peer-to-peer knowledge exchange and improvement of the quality of research reducing the effort and eliminating duplicated results;
- The connection between market and innovation actions becomes faster and seamless thanks to information flow; and
- Transparency during research is ensured while facilitating the progress and dissemination of ideas.

Data availability can be divided into two categories:
- Open Data: Data that is publicly shared for re-use and exploitation
- Private Data: Data that is retained by individual partners for their processes and tests
- Shared datasets will be allowed between consortium members to fulfil the project's objectives.

Most data will be embedded in the deliverables which will be publicly available on the TRANSFORMER Hub once they are accepted by CINEA. Additional data, namely collected tools and guidance will also be added to the Knowledge Hub that will be included as the functionality of the TRANSFORMER Hub, serving as a tool for two-way communication between Super-Lab stakeholders, the public and other target groups such as the media, educational institutions etc.

In addition, some of the data will also be available on Emilia-Romagna Region's specific portal called OpenDataER – dati.emilia-romagna.it. The APIs (Application programming interfaces) of dati.emilia-romagna.it portal conforms to the CKAN model and provides communication with the Apache Lucene project's Solr research platform, used by the portal. Certain data from Western Macedonia will be openly available on request and access through CERTH data lake repository. Data streams will be formulated according to a standard ontology (i.e., SAREF, Bricks Schema) with custom extensions. After access has been granted to external users, the "download by dataset" option from the API would allow specific data versions to be retrieved. Access will be possible through user registration and access approval by the administrator.

As the project progresses and data is identified and collected, further information on making data openly accessible will be outlined in a subsequent version of the DMP.

Making data interoperable

Interoperable data and metadata are described in the FAIR principles as those that use a formal, accessible, shared, and broadly applicable language for knowledge representation. They use vocabularies which themselves follow the FAIR principles, and they include qualified references to other data or metadata. What this describes is semantic interoperability. In other words, the data are described using normative and community-recognized specifications, vocabularies and standards that determine the precise meaning of concepts and qualities that the data represent.

As the project progresses and data is identified and collected, further information on making data interoperable will be outlined in subsequent versions of the DMP. In specific, information on data and metadata vocabularies, standards or methodology to follow to facilitate interoperability and whether the project uses the standard vocabulary for all data types present to allow interdisciplinary interoperability.

In particular, Emilia-Romagna Region declares its willingness to set up communities or joint working groups (also in the context of projects or national and international actions) with other interested public administrations to share skills and experience in the field of reuse of public data and cooperate in the creation of tools (IT and non-IT) implementing this policy. The APIs (Application programming interfaces) of dati.emilia-romagna.it portal conforms to the CKAN model and provides communication with the Apache Lucene project's Solr research platform, used by the portal. Using the dati.emilia-romagna API it is possible to search for data or import all dati.emilia-romagna data into another portal. To use the dati.emilia-romagna API by another application, it is necessary to make an HTTP request and parse the response, which by default is in JSON format.

Increase data re-use

For data to be **Reusable**, the FAIR principles reassert the need for rich metadata and documentation that meets relevant community standards and provide information about provenance. This covers reporting how data was created (e.g., survey protocols, experimental processes, information about sensor calibration and location) and information about data reduction or transformation processes to make data more usable, understandable or 'science-ready'. Reusability also requires that the data be released with a 'clear and accessible data usage license': in other words, the conditions under which the data can be used should be transparent to both humans and machines.

The initial date of release for the data identified above will follow the timeline/schedule (Gantt chart) of the deliverables as defined in the DoA (Table 1). The data quality assurance will be done before the publication of the relevant data-related deliverables.

The data will be usable by third parties after the end of the project and most of the data will fall under the license of CC BY Creative Commons Attribution. Further clarification will be outlined in a subsequent version of the DMP.

An exception will apply to the TSL of Western Macedonia, however, where data access will be restricted but once granted users can use or reuse them. Data collected for the Western Macedonia user case will be available upon request and approved by CERTH and ANKO at least 3 years after the project end.

Chapter 4: Allocation of resources

The allocation of resources to make the data FAIR is entailed in the budget allocated for each partner, especially under TWE and CERTH budget for respectively setting up the website and the Knowledge Hub.

Chapter 5: Data security

The management of the project knowledge and IPRs are specified in the Consortium Agreement (CA) that is signed by partners. Its content reflects, in some cases, complements to the terms and conditions defined in the Commission Contractual Rules. More specifically, the CA covers topics such as Individual and Joint Ownership of the knowledge, Protection of knowledge, Publication of results, Use and dissemination of knowledge arising from the project, access rights, Open Source and Standards, etc.

A holistic security approach will be followed, in order to protect the pillars of information security (confidentiality, integrity, availability). The security approach will consist of a methodical assessment of security risks followed by their impact analysis. This analysis will be performed on the personal information and data processed by the proposed system, their flows and any risk associated with their processing.

Security measures may include secure protocols (HTTPS and SSL), login procedures, as well as protection against bots and other malicious attacks, for example, CAPTCHA technologies. Moreover, the demonstration pilot sites apply monitored and controlled procedures related to data collection, integrity and protection. The data protection and privacy of personal information will include protective measures against infiltration as well as physical protection of core parts of the systems and access control measures.

According to the regulations on personal data protection, Emilia-Romagna Region will promote, where possible, in order to facilitate administrative transparency and make as much data as possible reusable (net of regulatory constraints), the publication of aggregated and anonymized data at a level of detail such as to not allow any identification, even indirect, of the subjects to whom the data refer. The areas in which there are limitations to the action of open data by public companies
administrations can be divided into five categories:
1. Not possess ownership of the data;
2. Process data which are contrary to the rules on the processing of personal data;
3. Process data not for the institutional purpose for which they were acquired or with respect to the administrative competence of the institution;
4. Release open data that infringes the rights of third parties;
5. Breach special rules such as State secrecy, statistical confidentiality, industrial secrecy, intellectual property, and limits on the publication of environmental data.

The TRANSFORMER project has a secure Knowledge Hub on the website and a SharePoint storage system in place for document management, storage system, and file sharing among members of the consortium. The TRANSFORMER Knowledge Hub and SharePoint server goes through regular security updates and has real-time protection. Data security issues will be assessed before the publication of data in the deliverables.

Procedures for Personal Data Collection and Processing

During TRANSFORMER, the active involvement of the general public and citizen science will be highlighted in all project phases (i.e. starting from T2.2 with the identification of existing structures and practices of regional decision-making and result validation in WP5.

This project involves collecting data, including personal data and metadata. All processing of personal data will be conducted following the provisions of 1) the GDPR Regulation (EU) 2016/679)24, 2) the Universal Declaration of Human Rights and the Convention 108 for the Protection of Individuals concerning Automatic Processing of Personal Data, and 3) the national laws applying its provisions, including those governing the acquisition of valid consent.

TRANSFORMER will generate various data including quantitative and qualitative research data, research reports, deliverables, internal evaluation data, participant lists, dissemination material, meeting minutes, press releases, and agendas. Confidential data, as well as data about persons, will be treated as such and only stored via password-protected platforms. Data not meant for the public and internal data will be shared by the consortium via a password-protected Intranet (SharePoint).

For the collection and processing of personal data during any stakeholder consultation and/or use cases realisation that will take place throughout the TRANSFORMER project, the following procedures should be followed:

- The collection of personal data from potential stakeholders will be set voluntarily.

- Invitations to the potential stakeholders to participate in each of the project's consultation processes will be given by individual Consortium partners, linked third parties and/or the Advisory Group (AG) members through their own network of contacts; therefore, no contact data will be shared without consent.

- An Informed Consent Form explaining the use of the personal data will be available as part of the registration form and its approval will be necessary for completing the subscription to workshops and webinars.

- Any time after subscription, stakeholders reserve the right to terminate their communication and cooperation with the TRANSFORMER consortium. Moreover, they also reserve the right to access and modify their data.

- Information about the objectives of the project and details about the consultation processes will be provided in advance to all external stakeholders who are interested in contributing to them. In addition, they will receive:

  - a simple description of the project in written/oral form;

  - the project goals;

  - the planned project progress;

  - the relevant consultation procedures;

  - advice on unrestricted disclaimer rights on their agreement;

  - information about any face and voice data recordings and the exact reference to the consent form (if applicable).

  - In this document, the privacy statements for handling the collection, processing, storage and protection of personal data will be explained.

- When participating in any consultation process, the stakeholders can indicate whether their personal data (e.g., name, surname, name of organisation) may be published in the project's reports or on the project's communication materials.

- Personal contact data (e.g., email) will be kept internally within the TRANSFORMER consortium and will not be published or accessible to external organisations and/or individuals;

- No secondary use of personal data will be performed;

- Personal data collected in the EU will not be transferred to entities in non-EU countries;

- Stakeholders will be approached based on their professional capacity and may be contacted for consultation only;

- No scientific research, analysis or processing will be performed on the personal data collected;

Any personal data and technical feedback/input collected in the context of the TRANSFORMER project will be stored electronically (e.g. in form of a word or excel files) on the project's shared extranet, which will be password protected and accessible by the project's consortium and linked third parties only. These data will not be shared and cannot be accessed and copied by any external party.

Chapter 6: Ethics

The TRANSFORMER consortium will ensure that research follows ethical principles and existing legal frameworks. Especially when human participants are part of the process, ethical research ensures that fundamental human rights, such as the right to privacy, the protection of personal data, the physical and mental integrity of a person, the right to non-discrimination and to the need to ensure the protection of

the environment and high levels of human health protection are protected at any time. The only question in the ethics survey that for TRANSFORMER has been answered with a "yes" is the question of whether the activity involved the processing of personal data. This is first and foremost related to the online part of the dissemination and communication activities. The planned measures are further described and explained below.

As a part of the cookies installed on the TRANSFORMER website, we will be processing bulk user data. For this, we will use GA or Piwik and the website will be hosted in the EU and on European servers, the system will be integrated with the CAPTCHA and Mollom spam protection system.

Also, as within the project, we will be connecting to other databases and processing or geolocation and personal data, we will ensure that the connection to other databases is compliant with the EU GDPR (General Data Protection Regulation) rules. EU GDPR rules will also be followed concerning the project's social media channels and promotion activities, when setting up invitation lists for events that will be hosted by project partners or when inviting people to focus group meetings. Personal data collected within TRANSFORMER will neither be passed on to unauthorised persons nor will be used for anything other than the implementation of the project. Confidentiality is protected by various measures which ensure that personal data neither falls into the hands of third parties nor can be viewed by them – this includes that personal data will be anonymised, password-protected, and stored on encrypted servers. It should be mentioned that within TRANSFORMER, also online surveys will be conducted. However, they will be conducted anonymously, i.e., no personal data is being collected.

Equally, we will respect all **intellectual property rights**, when it comes to publishing external content from other initiatives, partners, projects, or any other third parties.

All details related the secure systems of the TRANSFORMER website are listed in the Communication and Dissemination Strategy.

Information about ethics and data security from the Knowledge Hub.

The data that will be transmitted to the Knowledge Hub hosted at CERTH server infrastructure will be already (pseudo)anonymized, as appropriate mechanisms (such as random unique identifiers rotation) will be applied in the back-end system that will be developed under a "security by design" methodological approach. The data will never be used or shared for anything beyond the specific TRANSFORMER research purposes and any individual's personal information will never be processed without the data subject's explicit consent. Additionally, data subjects can request at any time a copy of their data, correct any data associated with them or request to erase their data completely and permanently from the Knowledge Hub.

Security Measurements

CERTH implemented technical and organisational measures to ensure an appropriate level of protection against the risks arising from processing and storage, such as accidental or unlawful destruction, loss, alteration, unauthorized disclosure, or access. The computer systems (servers) that will host the back-office of Knowledge Hub are parts of HIT-CERTH's server domain which is governed by several internal policies setting out data security and integrity. In this context, the following measurements will be taken:

1. **Distinct credentials for each application and user will be set** to ensure the security of personal data, including preventing unauthorized access to or use of personal data and the equipment used for the processing.

2. **Credentials on application configuration files will be encrypted** to avoid risks that are presented by processing, in particular from an accidental or unlawful destruction, loss, alteration, and/or unauthorized disclosure of that data. In this case, the protection is at the level of the physical device—and prevents the risk of compromising the storage itself, for example via copying the physical data out to another server.

3. **Optimal security configuration and firewall access will be set up.** The complete server infrastructure is being protected by a properly configured modern network firewall, set up with proactive measurements in order to prevent unauthorized access to the internal network systems. Pro-active measurements include –and are not limited to– the usage of dynamic access block lists updated regularly from trustworthy sources to include almost real-time threats, geo-blocking IP lists to deny access to geographical areas that are not in the scope of the project and network protocol/URL filtering.

4. **TLS encryption (HTTPS) will be enabled (data protection "on the move")**. The security measures taken include the usage of the TLS 1.2 encryption protocol encryption with 256-bit on every aspect of communication between external entities (back-end) and the back-office. Protecting data during transmission is necessary in order to avoid possible leakage and minimize relevant risks.

5. **Databases will be encrypted (data protection "at rest")** In this case, the protection is at the level of the physical device—and prevents the risk of compromising the storage itself, for example via copying the physical data out to another server.

6. **Every access attempt** will be **audited,** maintaining a record of processing activities under the responsibility.

7. **Verbose logs of data processing will be kept in .**txt format and/or other methods.

8. **Backup mechanism:** Automatic backup schedules will be set up to avoid data loss in case of a sudden hardware failure and/or system security breach.

Security incident response

As soon as a theft, data breach or exposure is identified, the process of removing access to that resource will begin. Furthermore, an incident response team will be assigned the responsibility to take necessary corrective action to remediate the incident. Such action shall include all necessary steps to preserve evidence in order to facilitate the discovery, investigation, and prosecution of crimes against IT resources.

The incident response team may follow the following guidelines:

1. Immediate disable WAN and LAN connectivity of the machine(s) where the security incident has been detected. In cases where such detection is not trivial, connectivity of the entire network infrastructure should be disabled until the affected and unaffected machines are confidently identified. This should prevent further exposure to the risk and will enable the team to initiate the process of determining the risk's nature and to take the appropriate steps to properly respond to the incident.

2. Identify the type of data involved in the incident as well as the volume of data exposed (if any). In case the nature of the incident requires it, the procedure provided for by the relevant GDPR regulations is initiated.

3. A thorough examination of the system and application logs as well as of additional existing evidence in order to identify the exact means by which the breach took place. If possible, remediation of the problematic point (emergency system and/or application update) is carried out on the latest available system backup which does not show any signs of a breach.

4. In coordination with the project's partners and the GDPR procedures, the decision on whether and when the system should be back online is being taken. For this decision to be taken, the incident response team provides all the available technical information to the interested parties and follows the instructions of the legal department.

Moreover, the TRANSFORMER consortium will ensure that all project activities including human beings are designed in a way that they comply with legal standards such as the regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 as well as with the rules of good scientific practice, such as the European Code of Conduct for Research Integrity, the OECD's Best Practices for Ensuring Scientific Integrity, the European RESPECT Code of Practice, and Data management Regulation (EU) 2016/679. Furthermore, the ethical standards and guidelines of Horizon Europe will be rigorously applied, regardless of the country in which the research is carried out. Further to the Horizon Europe ethical principles set out in Article 19 of Regulation (EU) No 2021/695, applicable national, international and organisation/ company measures – whichever is stricter – will be followed by the project partners.

Chapter 7: Conclusions

The TRANSFORMER DMP is based on the datasets for procedures and infrastructure that are anticipated at this point in the project. The first version of this deliverable aims at providing the first identified datasets and outlining a draft version of the components' specifications.

The next actions will be to focus on semantics and further clarification of procedures, participant and stakeholder engagement and identifying areas that need special attention. Changes to the datasets may occur after systematic studies on TRANSFORMER Super-Labs.

An important conclusion is that all the partners are responsible for different kinds of datasets. At this phase of the project, there are some difficulties to specify exactly all the relevant datasets for the projects' activities.

Chapter 8: References

EU Grants: Data Management Template (HE): V1.0

European Commission (2018), Turning FAIR into reality

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016:

https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG

Annexes

WP2

Database of TSL predecessors

DS_02_ Database_of_TSL_predecessors	
Data Identification	
Dataset Reference / Name	Database of TSL predecessors
Dataset Description	Excel database of existing predecessors of TRANSFORMER Super-Labs in the EU and beyond. The dataset consists of basic data about the project/structures (name, short description, sources etc.).
Source (e.g. which device?)	Excel (.xlxs)
Relevant Architectural Component	
Partners services and responsibilities	
Partner in charge of the data collection (if different)	RUB
Partner in charge of the data analysis (if different)	RUB
Partner in charge of the data storage (if different)	RC, CERTH
WPs and tasks	WP 2, Task 2.3
Metadata and Expected Size	
Info about metadata (Production and storage dates, places) and documentation?	- Data collection started in September 2022 and will end in November 2022 - the data is collected in an Excel sheet (stored at personal desktop PC and on SharePoint) - there is additional documentation (methodology of data collection is described in Deliverable D2.1: methodology)
Estimated volume of data	<100 KB
Data exploitation	
Data exploitation (purpose/use of the data analysis)	The database is part of Task 2.3 and basis for D2.1: The main objective of the Deliverable is to identify TSL predecessors in order to evaluate, assess and compare their conceptual framework and methodological approach for accelerating the transition towards climate neutrality. These findings are going to contribute to the development of the TSL concept and are a prerequisite for a common understanding of the

methodological approach and the conceptual grounding of evaluation and impact assessment within the TRANSFORMER project.

RUB - Database of regions

DS_02_ Database_of_regions	
Data Identification	
Dataset Reference / Name	Database of regions across Europe with potential and special need for the TSL approach
Dataset Description	Excel database of European NUTS2 regions (where data is available at this scale) based on statistical resources from esp. Eurostat and national offices on variables like emissions/CO2 footprint, energy intensity and mix, and industry structure.
Source (e.g. which device?)	Excel (.xlxs)
Relevant Architectural Component	
Partners services and responsibilities	
Partner in charge of the data collection (if different)	RUB
Partner in charge of the data analysis (if different)	RUB
Partner in charge of the data storage (if different)	RC, CERTH
WPs and tasks	WP 2, Task 2.1
Metadata and Expected Size	
Info about metadata (Production and storage dates, places) and documentation?	- Data collection will start in November 2022 and will end in January 2022 - the data is collected in an Excel sheet (stored at personal desktop PC and on SharePoint) - there is additional documentation (methodology of data collection is described in Deliverable D2.2: methodology)
Estimated volume of data	<1 MB
Data exploitation	
Data exploitation (purpose/use of the data analysis)	The database is part of Task 2.1 and basis for D2.2: The main objective of the Deliverable is to identify EU regions that would benefit the most from the TRANSFORMER Super-Lab

	approach (which focuses especially on innovation/leverage points at the intersection of socio-technical regimes in regions). For this, RUB will develop and apply a methodology to define the transition needs of all European NUTS2 regions (where data is available at this scale) based on statistical resources from esp. Eurostat and national offices on variables like emissions/CO2 footprint, energy intensity and mix, and industry structure. Based on this transition-need mapping, RUB will develop a typology of regions with the potential to become future TRANSFORMER Super-Lab regions.

Summary of expert interviews ad document analysis

DS_02_ Summary_of_expert_interviews_and_document analysis	
Data Identification	
Dataset Reference / Name	Summary of expert interviews & document analysis
Dataset Description	Summary of expert interviews and document analysis for the feasibility studies (mainly qualitative data, supplemented with quantitative data [key statistics] about the region).
Source (e.g. which device?)	Word (.docx)
Relevant Architectural Component	
Partners services and responsibilities	
Partner in charge of the data collection (if different)	RUB
Partner in charge of the data analysis (if different)	RUB
Partner in charge of the data storage (if different)	RC, CERTH
WPs and tasks	WP 2, Task 2.2
Metadata and Expected Size	
Info about metadata (Production and storage dates, places) and documentation?	- Data collection will start in November 2022 and will end in May 2022 - the data is collected in Word files (stored on personal desktop PC and on SharePoint)

	- there is additional documentation (methodology of data collection is described in Deliverable D2.3: methodology)
Estimated volume of data	<10 MB
Data exploitation	
Data exploitation (purpose/use of the data analysis)	The summary of expert interviews & document analysis is part of Task 2.2 and D2.3: For the feasibility studies, RUB will develop an appropriate mix of methods (e.g. focus groups, expert interviews, document analysis) to generate data about the Super-Labs and apply this methodology in tandem with the task partners. Firstly, a bottom-up mapping of the socio-technical regimes and their potential portfolio of solutions for climate transition in the TRANSFORMER regions will be conducted, and secondly, existing structures and practices of regional decision-making, including regime-actors, networks, strategies and policies will be mapped. Following a harmonised approach that allows for interregional comparisons, this data will be integrated into the form of regional SWOT analyses and made available to the stakeholders in the TRANSFORMER Super-Labs. Important aspects will be risks and benefits of different solutions, financial feasibility, key transformation challenges, innovative potential, available strategies, potential networks, social implications, and aspects of diversity (including gender).

ITL & RER

DS_03_ITL and RER	
Data Identification	
Dataset Reference / Name	To be defined
Dataset Description	To be defined
Source (e.g. which device?)	The data will be available from various sources depending on the necessary data
Relevant Architectural Component	To be defined
Partners services and responsibilities	
Partner in charge of the data collection (if different)	Regione Emilia-Romagna

Partner in charge of the data analysis (if different)	Regione Emilia-Romagna and Fondazione ITL
Partner in charge of the data storage (if different)	Regione Emilia-Romagna
WPs and tasks	WP2: Task 2.2, 2.3;
Metadata and Expected Size	
Info about metadata (Production and storage dates, places) and documentation?	For the general descriptive metadata Emilia-Romagna Region uses the national profile DCAT-AP_IT, according to the indications of the national guidelines. It is also possible to integrate the metadata provided by the DCAT-AP_IT model with additional metadata, according to each own necessity albeit in full compliance with the rules as defined in the DCAT AP_IT specification. The data will be made available in machine-readable formats.
Estimated volume of data	To be defined
Data exploitation	
Data exploitation (purpose/use of the data analysis)	The purpose of data collection and analysis is to understand the existing solutions and trends in the mobility and energy sectors and how they can be improved, harmonized and co-created by bringing together different stakeholders (e.g. regional and local authorities, regional agencies, clusters, etc.). The development of this participatory process will ensure that innovative solutions answer local needs and demands.

WP3

BMR

DS_01_ Regional initiatives as potential pilot cases	
Data Identification	
Dataset Reference / Name	Lists of initiatives as potential pilot cases from the four regions
Dataset Description	The four regions are expected to deliver lists of initiatives from the respective fields (leverage points for achieving climate neutrality)
Source (e.g. which device?)	
Relevant Architectural Component	

Partners services and responsibilities	
Partner in charge of the data collection (if different)	BMR
Partner in charge of the data analysis (if different)	BMR
Partner in charge of the data storage (if different)	CERTH, RC
WPs and tasks	WP3, Task 3.2
Metadata and Expected Size	
Info about metadata (Production and storage dates, places) and documentation?	- Data collection will take place from February to April 2023 - the data will be collected in an Excel sheet (stored at personal desktop PC and on SharePoint) - there is no additional documentation
Estimated volume of data	TBD
Data exploitation	
Data exploitation (purpose/use of the data analysis)	The database is not explicitly part of Task 3.2, but it can be an important basis for the deliverable 3.2: The main objective of the deliverable is to define Transition Super-Lab use cases (pilot cases)

ENoLL

DS_03_ENoLL	
Data Identification	
Dataset Reference / Name	Living Lab internal and external stakeholders
Dataset Description	We will be gathering data on the internal stakeholders (which includes names of people who are working in the Living Labs) and information on external stakeholders who are part of the quadruple helix. In the latter case no personal data will be collected.
Source (e.g. which device?)	The data gathering workshop will take place on a Miro board and will be translated to a document stored on Sharepoint
Relevant Architectural Component	TBD
Partners services and responsibilities	
Partner in charge of the data collection (if different)	/
Partner in charge of the data analysis (if different)	/

Partner in charge of the data storage (if different)	ENoLL and RC
WPs and tasks	WP3 and WP6, more specifically T3.2 and T6.4
Metadata and Expected Size	
Info about metadata (Production and storage dates, places) and documentation?	Storage on SharePoint of RC
Estimated volume of data	To be defined
Data exploitation	
Data exploitation (purpose/use of the data analysis)	The data will be used for preparation of the capacity building program/training of the partners

ITL & RER

DS_03_ITL and RER	
Data Identification	
Dataset Reference / Name	To be defined
Dataset Description	To be defined
Source (e.g. which device?)	The data will be available from various sources depending on the necessary data
Relevant Architectural Component	To be defined
Partners services and responsibilities	
Partner in charge of the data collection (if different)	Regione Emilia-Romagna
Partner in charge of the data analysis (if different)	Regione Emilia-Romagna and Fondazione ITL
Partner in charge of the data storage (if different)	Regione Emilia-Romagna
WPs and tasks	WP2: Task 2.2, 2.3; WP3: Task 3.1, 3.2, 3.3, 3.4; WP4: Task 4.1, 4.3; WP5: Task 5.2, 5.3; WP6: Task 6.1, 6.2, 6.3, 6.4
Metadata and Expected Size	
Info about metadata (Production and storage dates, places) and documentation?	For the general descriptive metadata Emilia-Romagna Region uses the national profile DCAT-AP_IT, according to the indications of the national guidelines. It is also possible to integrate the metadata provided by the DCAT-AP_IT model with additional metadata, according to each own necessity albeit in full compliance with the rules as defined

	in the DCAT AP_IT specification. The data will be made available in machine readable formats.
Estimated volume of data	To be defined
Data exploitation	
Data exploitation (purpose/use of the data analysis)	The purpose of data collection and analysis is to understand the existing solutions and trends in the mobility and energy sectors and how they can be improved, harmonized and co-created by bringing together different stakeholders (e.g. regional and local authorities, regional agencies, clusters, etc.). The development of this participatory process will ensure that innovative solutions answer local needs and demands.

CERTH & ANKO

DS_03_CERTH AND ANKO	
Data Identification	
Dataset Reference / Name	To be defined
Dataset Description	Examples of the datasets to be used: Mobility patterns data, raw trajectories, road network, public transport infrastructure (routes, vehicles of PuT fleet, stops etc), PuT demand volume, sociodemographic data, PuT estimated energy demand Volume of energy produced by RES in the WMR/ Energy supply data Locations of RES and H2 production Number and locations of charging points for electric vehicles Locations of potential storage facilities CO2 emissions in agriculture
Source (e.g. which device?)	Various sources depending on the required data
Relevant Architectural Component	To be defined
Partners services and responsibilities	
Partner in charge of the data collection (if different)	ANKO
Partner in charge of the data analysis (if different)	CERTH

Partner in charge of the data storage (if different)	CERTH, RC
WPs and tasks	The data are going to be collected and utilized within activities of WP3 and WP4 during the use cases implementations and evaluation. Further analyses may be needed during the impact and tools/structures assessment of TSLs pilots in WP5.
Metadata and Expected Size	
Info about metadata (Production and storage dates, places) and documentation?	Dataset will be accompanied by detailed documentation of its contents. The metadata files may include (a) a description of the pilot setup and different experiment parameters, methods followed to generate the dataset, (b) aggregated data in terms of averages (e.g, mean trips per day per household, kilometres travelled, peak hours, etc.)
Estimated volume of data	To be defined
Data exploitation	
Data exploitation (purpose/use of the data analysis)	The data will be collected and analysed in order to acquire a complete overview of the current situation in mobility, energy, environment, agriculture and circular economy and the trends for each of the abovementioned sectors. This will provide valuable insights on the gaps and the opportunities enabling the efficient implementation of the use cases during the project.

WP4

RC

DS_04_RC Roadmap	
Data Identification	
Dataset Reference / Name	Roadmap data
Dataset Description	This dataset will be developed for the development of the TSL Roadmap, identifying patterns and lessons learned from the TSL development

Source (e.g. which device?)	The data used will be the ones collected in WP2 and WP3 and additional data may be collected through stakeholders' interviews and surveys
Relevant Architectural Component	
Partners services and responsibilities	
Partner in charge of the data collection (if different)	RC
Partner in charge of the data analysis (if different)	RC
Partner in charge of the data storage (if different)	RC
WPs and tasks	WP4, T4.1
Metadata and Expected Size	
Info about metadata (Production and storage dates, places) and documentation?	Data collected will be on the project SharePoint
Estimated volume of data	To be estimated at a later stage
Data exploitation	
Data exploitation (purpose/use of the data analysis)	The data will be used to complete T4.1