



TRANSCEND

Deliverable 6.2: DATA MANAGEMENT PLAN

Author(s): Sina Khatami, Giuliano Di Baldassarre, *Uppsala University*

With contributions from: C. Dionisio Pérez-Blanco, *Universidad de Salamanca*; Timothy Foster, *University of Manchester*; Hadi Jaafar, *American University of Beirut*

Version	Changes	Date	By
v1	Original submission	30/06/2023	Sina Khatami, Giuliano Di Baldassarre, <i>Uppsala University</i>
v2	1. Table of contents for figures and tables added 2. Formatting issues addressed (page numbers added) 3. Broken links have been fixed 4. Section 3 added 5. Annex xxxx added 6. Proofreading	17/10/2024	Giuliano Di Baldassarre, Omar Cenobio Cruz, <i>Uppsala University</i> ; Héctor González-López, <i>Universidad de Salamanca</i>

Deliverable 6.2 – Data Management Plan

Title	DATA MANAGEMENT PLAN
Author(s)	SINA KHATAMI, GIULIANO DI BALDASSARRE
Organization(s)	UPPSALA UNIVERSITY
Reviewer(s)	TIMOTHY FOSTER (THE UNIVERSITY OF MANCHESTER), HADI JAAFAR (AMERICAN UNIVERSITY OF BEIRUT), C. DIONISIO PÉREZ-BLANCO (UNIVERSIDAD DE SALAMANCA)
Deliverable number	6.2
Submission date	30/06/2023

Prepared under contract from the European Commission

Grant Agreement no. 101084110

This publication reflects only the authors' views and the European Commission is not liable for any use that may be made of the information contained therein.

Start of the project:	01/01/2023
Duration:	48 months
Project coordinator organization:	Universidad de Salamanca
Related Work Package:	6
Type of Deliverable:	Report
Due date of deliverable:	June 2023
Actual submission date:	30/06/2023

Dissemination level

- ☒ PU = Public, fully open, e.g. web
- ☐ CO = Confidential, restricted under conditions set out in Model Grant Agreement
- ☐ CI = Classified, information as referred to in Commission Decision 2001/844/EC.



The TRANSCEND Project has received funding from the European Union's Horizon Europe Research and Innovation Programme under the Grant Agreement 101084110

Executive summary

Data management plan (DMP) is a formal document that outlines how data are to be handled with respect to the FAIR (Findable, Accessible, Interoperable and Re-usable) data principles and in line with the “Open research data pilot” EU initiative. The goal of a DMP is to present and ensure partners observe all relevant aspects of data management, metadata generation and data preservation from the beginning of the project to ensure that data are well-managed both during the project’s course and beyond.

This deliverable provides the first version of the DMP for the TRASCEND project, 6 months from the start of the project. To prepare the current version of the document a number of reference data management documentation have been consulted by the authors and reviewed to produce a best-practice, self-consistent plan.

The document outlines how the research data collected or generated during the project will be handled by the project Beneficiaries during and after the TRASCEND project. It describes which standards and methodology for data collection and generation will be followed, and whether and how data will be shared and preserved.

TRASCEND will use the Zenodo open research data repository, to provide open access to all publications from the project, as well as the underlying datasets (with all datasets based on personal data irreversibly anonymised). Data will be uploaded to Zenodo within the European Commission Funded Research (OpenAIRE) Community, as well as a community specifically created for the TRASCEND project. Uploaded datasets will be assigned a Digital Object Identifier (DOI) to ensure it can be properly referenced, and provided with the relevant metadata that includes the grant agreement number.

The DMP is intended to be a living document in which information will be updated, if necessary, for example to cover changes in available storage facilities or address new relevant data management aspects that arise during the TRASCEND project. The DMP aims to maximize the utility and value of the project’s data and knowledge for other research projects focused on water themes, as well as for stakeholders who have a vested interest in the outcomes of the project.

Contents

1. Data summary	6
1.1. Data size	8
1.2. Data utility	9
1.3. DMP update and revision	9
2. FAIR data	9
2.1. Making data findable, including provisions for metadata	9
2.1.1. Naming conventions	10
2.2. Making data accessible	10
2.2.1. Open Access publication	11
2.3. Making data interoperable	12
2.3.1. Data Exchange	13
2.3.2. Data Vocabularies for Interoperability	13
2.4. Increase data re-use	13
2.4.1. Data Quality Assurance	13
3. Data management	14
4. Allocation of resources	15
5. Data security	15
6. Ethics	15
7. Other issues	16
Annex I Data file formats	17
Annex II Example of data and level of confidentiality for the Timpaki LL	18

List of Tables

Table 1. Open Access plan of the academic partners	12
--	----

List of Acronyms

Acronym	Meaning
CMIP6	Coupled Model Intercomparison Project phase 6
DMP	Data Management Plan
DOI	Digital Object Identifier
EAB	External Advisory Board
FAIR	Findable, Accessible, Interoperable, Re-usable
ISIMIP	The Inter-Sectoral Impact Model Intercomparison Project
TAP	Transformational Adaptation Policies
TRANSCEND	Transformational and Robust AdaptatioN to water Scarcity and ClimatE chaNge under Deep uncertainty
WPs	Work Packages

1. Data summary

TRANSCEND (<https://transcend.usal.es/>) is an interdisciplinary research project that aims to identify and catalyze the adoption of Transformational Adaptation Policies (TAP) to water scarcity, including innovative allocation systems and economic instruments, that are robust and adaptable to uncertainty and change, while simultaneously achieving equitable and sustainable economic growth and welfare. To this end, TRANSCEND will develop a groundbreaking ecosystem of innovation that combines three key pillars: (1) a knowledge network for stakeholder engagement and knowledge sharing, (2) an actionable modeling suite that integrates interdisciplinary socio-ecological science and ensemble forecasting to guide TAP design, and (3) an accounting and monitoring toolbox that supports implementation and enforcement of TAP in practice.

The datasets, metadata, and workflows generated by TRANSCEND will be made **Findable, Accessible, Interoperable and Reusable (FAIR)** for other research projects and researchers, stakeholders, and the civil society. The overarching domain of the data that will be collected or generated within our project is “**geospatial data/models of hydro-climate and agro-economics**”, which include:

- A.** policy datasets documenting learning, both from Living Labs and beyond, about the current status of socio-ecological and river basin modeling (WP3), water policy monitoring and enforcement (WP4), and Transformational Adaptation Policies (TAP) governance and financing (WP5).
- B.** original water accounting datasets, including information about different aspects of water supply, availability, and/or usage within living lab catchments, to be generated through earth observation and ground-truth datasets (WP4) to support TAP design and implementation.
- C.** geospatial datasets that provide inputs to support TAP modeling and monitoring (WP3 and WP4) developed through the harmonization and merging of existing secondary climate, hydrologic, agronomic, microeconomic and macroeconomic datasets (WP4).
- D.** simulation datasets using the modelling suite, including a groundbreaking database of plausible futures (D3.3) that complements major ongoing efforts in the international (WP3).
- E.** research reports and peer-reviewed journal papers (all WPs).

A are primary data collected through a variety of methodologies, including observational data, questionnaires (quantitative datasets), and literature reviews — acquired both for the project’s Living Labs and for other catchments globally.

Purpose of data collection/generation: These data are being generated through various reviews and questionnaires as part of the white papers. Questionnaires from stakeholders may serve to validate data.

Main sources for data collection/generation: Data will be collected through primary surveys and questionnaires in Living Labs and other river basins globally, along with information gathered through synthesis of secondary literature and policy documents in the public domain.

Data format: The data will be documented in excel sheets or text files, with accompanying detail in project white papers. Datasets will be made available at different times during the project.

B are primary data developed through use of Earth Observation imagery and other geospatial data, which characterize different aspects of water supply, availability, demand/use that are of relevance to stakeholder modeling and monitoring of TAP. These datasets will be primarily developed for the project's Living Lab catchments, utilising ground validation data provided by the corresponding water/catchment authorities and agencies of a given Living Lab. Final validated data products will though cover larger geospatial domains, or have the potential to be extended using supporting models codes. Water Accounting data inputs may include observed flows, water storage change, topography, land cover classification, precipitation, and evapotranspiration data, among others.

Purpose of data collection/generation: These data are generated in terms of monitoring products. The availability of accurate and up-to-date water accounting data facilitates better monitoring of water resources. These data are essential for establishing water use limits (as baseline or benchmark) and for assessing deviations from this baseline under different simulations scenarios.

Main sources for data collection/generation: All data will be sourced from open access datasets and databases hosting satellite and other geospatial data, such as Google Earth Engine, Copernicus, and NASA Earth Data.

Data format: Data will be provided in multiple formats according to product type and user needs, with common formats including raster images, NetCDF files, text files, and accompanying R/Python/MATLAB scripts. Datasets will be made available at different times during the project, including through different offline and online web-based tools.

C are secondary datasets that are based on processing and coalescing existing datasets of hydro-climate, ergonomics, and micro-/macroeconomics. The hydro-climate data include geospatial datasets that are remotely sensed, based on ground measurements, and/or models.

Purpose of data collection/generation: These data are the required inputs for developing the simulation models and their corresponding scenarios.

Main sources for data collection/generation: All data are sourced from open access and available online as well as the relevant agencies of a given Living Lab; such as the European Climate Assessment & Dataset (e.g. ensemble of gridded daily hydroclimate observational datasets <https://www.ecad.eu/download/ensembles/download.php>), [ISIMIP](#) and [CMIP](#), datasets for the regionalization of macroeconomic models, or land use data and scenarios for policy design and assessment of responses by microeconomic models such as [LUMIP](#). The sources include official country and city level statistics and reports, WHO data, census and population data, regional and household statistics and questionnaires undertaken by private, public, or non-governmental organisations. These may be available online and/or only accessible on-site (hard, unpublished reports).

Data format: These data will be in multiple formats including (but not limited to) excel sheets, text files, and NetCDF. We will process the data and develop workflow so that all data are integrated and ready to use regardless of their format for a given model or analysis.

D include the simulation outputs that will be post-processed to be compatible with the pre-processed input data (C). They also include ground-breaking plausible socio-environmental scenarios/futures for the Living Labs to explore TAP. Scenarios will be based on theoretical insights from critical social sciences (e.g. <https://www.nature.com/articles/s41558-022-01546-8>), to be simple, flexible and easily adaptable to different types of models/cases. Data products of interest relates to; future scenarios, human-water interactions (e.g. population maps, urbanisation patterns, water infrastructure), hydro-climate including past events of hydrological extremes (droughts and floods), agro-economic (e.g. crop pattern, yield, water footprint), and vulnerability aspects such as poverty. The new datasets will be made available with corresponding metadata in xml format, and provided in commonly used geospatial data formats such as GRID files, shapefiles and NetCDF files. Some analyses might be conducted through cloud computation services due to large data sizes, this will be made available through sharing the processing codes.

Purpose of data collection/generation: The model simulations will be used to explore the impacts of policies and identify robust strategies given different scenarios.

Main sources for data collection/generation: These data are generated by the simulation models that are used/developed in this project.

Data format: These data will be in multiple formats including (but not limited to) excel sheets, text files, and NetCDF. Simulation models (D3.3) will be provided in text file format, ensuring compatibility and ease of use. Additionally, scenario databases will be obtained and disseminated in multiple interoperable formats, including database formats (such as dbf, mdb, csv), NetCDF, and spatial raster formats (such as tiff). Further description of file formats are provided in the Annex I. This approach enables flexibility and accessibility for users working with different software tools and platforms. We will process the data into a unified format (text files and NetCDF) that are easy to store, publish, and/or process for further analyses.

E are peer-reviewed papers and reports from literature reviews, syntheses, and analyses published in open access journals, including the description of the data, sources, and workflows used to ensure repeatability and reproducibility.

1.1.Data size

We anticipate that the entire dataset of the project will be approximately 1 TB. Datasets will be made available during and/or at the end of the project, depending on the publications — i.e. they will be made available at the time of publication or pre-publication — and on the embargo periods and access procedures. The models, scenarios, monitoring tools and datasets developed in the research project which will be published and made available on GitHub, linked to Zenodo where we publish the raw data. We note that it may not be possible to provide open access of some sensitive or copyrighted data. Published papers drawing from sensitive datasets will be referenced in the metadata of surveys and of other datasets produced by the project to ensure future data users have a reference to the datasets that will not be available. Following the EU guidelines regarding on Data Management Plans, this plan might be updated - if needed - during the project lifetime.

1.2.Data utility

The primary target group of the data produced within the project are the Consortium Members (CMs) including the partners of each Living Lab, as organised storage and documentation of data will be key to support consortium-wide collaboration and scientific and training interactions. To enhance the utility and re-usability of the data, the DMP aims to provide access to raw data, metadata, methodological data, and source code scripts that facilitate running potential models. In cases where data coding is involved, comprehensive code documentation in the form of manuals or tutorial videos will be provided, enabling partners of each Living Lab as well as the broader community to effectively run the various models and replicate the analyses. The secondary target group is the broader community including academia and research institutions, and public and private sector practitioners. Here, the key aspect will be ensuring the reproducibility of results and adherence to the FAIR principles.

1.3.DMP update and revision

DMP is a living document: we emphasise that DMP is a living document and will likely be updated when adjustments are necessary (i.e. if there are changes in data storage or preservation requirements; if there are updates in the data collection methods or tools being used in the project; if there are changes in the policies or regulations regarding data sharing or access; and if stakeholders provide feedback or recommendations that require modifications to the data management plan). The project consortium will employ periodic updates and versioning in the DMP document, by revisiting and disseminating it annually, in-line with the periodic evaluation of the project, to address the inclusion of new datasets, output items, and other possible reforms.

2. FAIR data

2.1.Making data findable, including provisions for metadata

The project will ensure that both publications and all raw datasets that are not subject to specific restrictions will be made accessible and that enough information is provided to ensure that researchers are able to read, use and interpret the data in the future. First, all publications will be open access and will have a DOI or a similar persistent identifier. Further, information on the repository used for the original datasets will be provided in papers and vice versa. Concurrently, for each dataset collected, processed and/or generated in the project, the project team will develop several strategies to ensure they are deposited Open Access in a repository and effectively described with appropriate metadata. The metadata will provide information on the following:

- **the domain and possible uses** - this set of metadata focuses on the nature of the dataset (e.g. type of data, data source); the scale of the dataset (e.g. local, regional, national, global); to whom could the dataset be useful (e.g. scholars or practitioners of water resources management); scientific publication associated to a given dataset; any restrictions that may apply; information (if applicable) on the existence of similar datasets; possibilities for integration with other datasets collected in the project and potential for reuse.

- **methodology and research design** - this set of metadata encompasses study design, workflows, and all other information necessary for a secondary analyst to use the data accurately and effectively.
- **organising principles**: each dataset will also outline the file-naming convention used and the protocol used to structure datasets produced by the project.
- **geospatial metadata**: new geospatial datasets will be accompanied with XML files giving the metadata according to industry standards. The metadata will include information about the identification, constraint, extent, quality, spatial and temporal reference, distribution, and version of the dataset. We will follow the ISO 19115-1:2014 (Geographic information — Metadata) standard (<https://www.iso.org/standard/53798.html>), and the general documentation about the dataset will also be provided in a README file.

It will be the responsibility of each researcher to annotate data with metadata and to code their files according to a protocol agreed by the team. The project Coordinator is responsible for monitoring researchers' practices and ensure data is being properly processed, documented, and stored.

2.1.1. Naming conventions

Following a consistent and precise naming convention facilitates the process of dataset access and retrieval for the future scientific and broader community. TRANSCEND encourages the use of a standard naming convention given to all its public domain documents as follows:

TRANSCEND-WPX#-DOCKEYWORD-v#Date#

"TRANSCEND" stands for TRANSCEND project.

"WPX#" stands for the work package under which the data lies.

"DOCKEYWORD" indicates a keyword associated with the file that identifies it further. "-" is a short dash that indicates a separator between elements.

v#Date# indicates the file version based on the date of the last modification.

For example, this document is named **TRANSCEND-D6.2-DataManagementPlan-v2023-06-27**, where *D6.2* indicates the deliverable item, *DataManagementPlan* refers to the document keyword, and *v2023-06-27* is the last version of the file by date.

This naming convention may be revisited throughout the course of the project based on generated outputs.

2.2. Making data accessible

Data management and storage will be done according to the procedures set in place by the Horizon Europe (HORIZON) Programme Guide on FAIR Data Management. The plan will vary depending on the data sets and its origins.

Internal rules for implementing Open Science across all activities of the project will include:

- **Early and open sharing of our results**: All deliverables and relevant milestones, training material and conference contributions will be made publicly available as early as possible, including as drafts, pre-prints or registered reports (All WPs).

- **Early and open sharing of simulation results and databases:** All original (simulations in the database of plausible futures, monitoring and water accounting data) and secondary databases (built through the harmonization of primary databases) will be made publicly available as soon as all internal quality checks are finished in widely used open access portals such as [Aquastat](#), [Climate Adapt](#), and [Google Earth Engine](#). We will also engage with and share our data via more targeted science- ([ISIMIP](#), [CMIP6](#)) and industry-oriented data portals ([Water Risk Filter](#)) (WP3, 4 & 6).
- **Open review:** All deliverables will be subjected to open peer-review, with a focus on obtaining feedback from stakeholders in the living labs, in addition to internal review and quality assurance mechanism (All WPs).
- **Open co-creation:** Formats and agendas of local and international workshops and other science-policy events, as well as content of training courses and outreach events, will be co-created with stakeholders in the knowledge networks. All major events, including local and international workshops where scenathons will take place, webstivals and any relevant event informing the co-creation process, will be registered and the video made publicly available in the project website (WP1 & 2).
- **Gold open access:** We will ascribe to Gold Open Access option, leveraging on financial resources made available by our organizations in the context of the transformational agreements recently subscribed with major publishing houses (such as Elsevier, Springer, Wiley, ACS), which shift away from payment to read and towards payment to publish open access. Working papers will be also made available in online repositories to ensure wide and timely dissemination, verification, and immediate re-use of research results (All WPs). All key deliverables and exploitable results will be made public under gold open access rules and Creative Commons Attribution (CC BY) copyright license, for which anyone will be permitted to adapt, remix, distribute and build upon the project's results, including for commercial use, as long as the source is properly cited, in compliance with the EU copyright legislation.

Beyond informing co-creation processes in labs, we will work in partnership with [CMIP6](#) and [ISIMIP](#) to ensure wide dissemination of this data across scientists and decision-makers. Moreover, we will embed the database of plausible futures into the core functionality of the [Water Risk Filter](#) and other major industry platforms, so as to create tailored results that are user- and industry-relevant. These efforts will be instrumental towards achieving a lasting legacy of our results (input to T6.4). All datasets, except for data held sensitive under the [EU's data protection law](#), will be open access via [FAO Aquastat](#), [EEA's WISE](#) and [Climate Adapt](#) portal, after the completion of all internal quality checks.

2.2.1. Open Access publication

TRANSCEND partners are encouraged and supported through the resources made available by our organizations in the context of the transformational agreements recently subscribed with major publishing houses (such as Elsevier, Springer, Wiley, ACS) to publish their research outputs as Gold Open Access. Open access strategies of all academic partners are listed in the table below:

Table 1. Open Access plan of the academic partners

Academic Partner	Open Access Strategy
Universidad de Salamanca	have signed the transformative agreement with main publishing houses to shift from pay per read to pay per publish, thus making all publications immediately gold open access. More information: https://bibliotecas.usal.es/noticia/acuerdo-transformativo-con-la-editorial-elsevier-para-la-publicacion-en-abierto-de-articulos
Uppsala University	have signed the transformative agreement with main publishing houses to shift from pay per read to pay per publish, thus making all publications immediately gold open access. More information: https://www.ub.uu.se/publish/open-access/OAagreements/
Valencia Polytechnic University (Universidad Politécnica de Valencia)	have signed the transformative agreement with main publishing houses to shift from pay per read to pay per publish, thus making all publications immediately gold open access. More information: https://poliscience.blogs.upv.es/acceso-abierto/publicacion-en-acceso-abierto/acuerdos-editoriales/
The University of Manchester	1. As a minimum, a version (unformatted accepted paper) is deposited in UoM's institutional repository (Pure) to make each paper green OA 2. Funding is provided from university to make some (not all) articles gold access, and UoM also has transformative funding agreements with a number of publishing houses/journals. More information: https://www.library.manchester.ac.uk/services/research/open-research/access/funding/
American University of Beirut	AUB Libraries is participating in the membership program of main open access publishers, with waivers and discounts for university affiliates. More information can be found here.

2.3. Making data interoperable

Interpolability will be ensured by identifying metadata that best describe the datasets. This will vary across different parts of the dataset. Yet, as discussed earlier, each dataset will specify domain and possible uses, methodology, research design and organising principles. Concurrently, *Zenodo* also provides a standard classification that includes Title, Creator, Date, Subject, Description, Text explaining the content of the dataset and other contextual information, Details of the file format, Resource Type and Identifier (DOI).

To ensure re-combination of datasets within and beyond the project (i.e. datasets from different origins), metadata will also provide information on: (1) other datasets produced by the project, (2) information (if applicable) on the existence of similar datasets that were not produced by the project, as well as (3) possibilities for integration with other datasets collected in the project.

Last, to ensure that the new geospatial datasets are exchangeable and reusable by researchers from other countries, institutions and organisations, TRANSCEND will, as much as possible, adhere to standards for formats compliant with available (open) software applications. First, the project members will make sure that the geospatial datasets are shared in standard geospatial file formats (e.g. GRID files, shapefiles, NetCDF), which can be reused with a number of open access software such as QGIS and Panoply. Second, the corresponding geospatial metadata will follow industry standards for Geographic information metadata XML.

2.3.1. Data Exchange

Our strategy to exchange data is to:

- Within the Consortium: use OneDrive for generic and everyday sharing of files. OneDrive has been selected due to its compliance with the Regulation (EU) 2016/679 (General Data Protection Regulation). In addition, we will use other open access platforms for sharing codes and workflows, seminars, and access to infrastructures such as supercomputing or cloud computing services.
- Beyond the Consortium: use working papers and journal publications with standard vocabularies.

2.3.2. Data Vocabularies for Interoperability

Vocabulary used will be common and standardized. In case project-specific vocabularies were used, mappings/glossary of terms to more commonly used terms will be provided. Experts of the Consortium will thoroughly review data and metadata standards and will ensure the use of formats commonly accepted in the water and climate community.

2.4. Increase data re-use

All publications produced by the project will be open access. We do not expect to produce any patents and most datasets will become open access (with the exception of those that might be restricted, e.g., some confidential data shared by stakeholders for calibration of microeconomic models, which cannot be publicly made available). Papers will be made available under the Creative Commons CC0 license and attached CC-BY Attribution license. This allows other researchers to distribute, remix, and build upon our work while the creators are given recognition. Some new datasets might, however, need to follow other license agreements due to parent dataset license restrictions. Datasets will be made available at different times.

2.4.1. Data Quality Assurance

TRANSCEND Consortium will ensure that quality assurance processes are established through rigorous and continuous quality check and review of published data, including datasets, reports, source codes, and other publications.

TRANSCEND partners will publish research outputs in international journals with high impact factors, which implies continuous external review of data and model quality check.

TRANSCEND will use an internal review process for data quality assurance as follows:

data-related deliverables will be delivered at least 1.5 months ahead of the submission deadline to the coordinator, who will revise the document with the support of two other PIs not involved in the relevant task/deliverable.

When necessary, the WP leader can request support from other Consortium partners or the External Advisory Board (EAB) leveraging on the resources made available for external reviews in the budget.

Reviews will be sent (at least) one month before deadline back to deliverable authors, who will ensure the deliverable is uploaded before the deadline.

Internal and external (through peer-reviewed publications and other knowledge networks) will ensure high quality standards of TRANSCEND data.

3. Data management

The management of workflows, models, and simulation outputs are already included in the previous sections, as they are an integral part of the research output. TRANSCEND Consortium Members will ensure the establishment of quality assurance processes over research outputs by conducting thorough and ongoing checks and reviews of published data, including datasets, reports, source codes, and other publications.

The Work Package (WP) leader will play a crucial role in overseeing and managing the complex interconnections between various key aspects of data governance. This includes ensuring robust data security protocols are in place, defining and enforcing access levels to maintain appropriate accessibility, and managing intellectual property rights (IPR) to safeguard the ownership and usage of data. Additionally, the WP leader will oversee the implementation of embargo periods to control the timing of data release, ensuring compliance with both ethical considerations and legal regulations. Furthermore, the WP leader will be responsible for coordinating the secure and efficient storage of data, guaranteeing that all legal, ethical, and organizational standards are met. This comprehensive approach ensures that the project's data management practices are both rigorous and compliant with relevant frameworks and guidelines.

To ensure interoperability and long-term digital preservation, it is essential to maintain an up-to-date description of all types of data used and produced throughout the project. This description should include details such as the proposed file formats, software required to access and render the data, and any other technical specifications critical for future use. Deliverable 3.2 ("Modeling Suite Prototype and Sourcebook") provides a comprehensive overview of the data sets collected/to be collected at each living lab, outlining the scope and nature of the data involved. Annex II complements this by offering a detailed example from the Timpaki Living Lab (LL), where a list of datasets used by various models is provided, alongside their respective confidentiality levels (e.g., open access or restricted access). The project will adhere to Open and FAIR data principles, ensuring that data is shared transparently and responsibly, while respecting the confidentiality restrictions of each dataset. This structured approach will ensure that the data remains accessible, usable, and secure, both during and beyond the project's duration.

4. Allocation of resources

As a part of our communication and dissemination activities, we will produce scientific articles and special issues in JCR Q1 journals showcasing project results, including the publication of 10+ papers per year. We will submit and publish one paper showing the results of the TRANSCEND Project to Nature and/or Science. The articles in high impact journals will be accompanied by short webinars. All papers will be published gold open access leveraging on resources made available by our organizations, and working papers published in online repositories.

Publication in Gold Open Access is estimated as 2500€ per paper, with 10+ paper per year, it is estimated as 22,500€ per year. These costs are paid by the institutions of the academic partner that is leading the publication. If that institution has not signed a transformative agreement for open access publication, the cost will be covered by the following author ascribed to an institution for which there is an agreement. Partner institutions which have signed transformative agreements for open access publication are listed in

Table 1.

Data curation including managing data during the project and preparing them for the repository, will be done by the researchers collecting the data and, thus, will not lead to additional costs. The project Coordinator will oversee the process and ensure continuity of data monitoring at the end of the project. Zenodo is a free repository and, therefore, the project will not incur extra cost to store the datasets.

5. Data security

Archiving and long-term preservation of the data will be done by storing them on a password protected space provided by the project Coordinator (e.g. SUNET Box: cloud file storage and sharing). Data will be preserved in perpetuity to allow for the possibility of a longitudinal study, where interviewees have agreed to do so and where datasets are not subject to embargoes or other restrictions. In doing so, in addition to the data storage in repositories (as explained in earlier sections), The coordinator will store the data in a cloud drive of the university, and will make a monthly security copy of the data in the cloud drive in a NAS drive with 2 hard disk drives in parallel to ensure that in case of failure of the cloud drive or one of the hard disk drives, the other has the information stored.

Data that will be deemed reusable for e.g. comparative or integrative purposes will be processed (e.g. converted to an easily accessible and reusable format) by the researcher responsible for data collection. As mentioned earlier, publication of some datasets might be restricted due to privacy and ethical concerns.

6. Ethics

We do not expect any ethical or legal issues for the sharing or storing the datasets, other than the possible restrictions of some of the sensitive data that we acquire from the Living Lab authorities/partners. In any such cases, the ethical or legal aspects will be considered carefully following the EU's rules and regulations, and will be acknowledged accordingly in any related reports or papers. An *ad-hoc* deliverable (Ethical Management Plan, D7.2) has been prepared to address all relevant ethical issues arising during the project.

7. Other issues

There are no other issues to be mentioned at this stage, further to what are already discussed. We will update this document throughout the project and make notes of any possible matters that may arise throughout the project lifespan.

Annex I Data file formats

- DBF is a file format commonly used for organizing structured data in a tabular format. It stands for "dBase File" and was initially developed for the dBase database management system. DBF files store data in a simple columnar structure, with each column representing a different field or attribute. They can be easily opened and edited using spreadsheet software or database management systems.
- MDB is a file format associated with Microsoft Access, a popular relational database management system. MDB stands for "Microsoft Database" and is primarily used to store and manage structured data. MDB files can contain multiple tables, each with its own set of fields and records. They support more advanced data management capabilities, including relationships between tables, queries, forms, and reports.
- CSV is a plain text file format used to store tabular data in a simple, human-readable format. CSV stands for "Comma-Separated Values," although other delimiters like semicolons or tabs can also be used. In a CSV file, each line represents a row of data, and the values within each line are separated by the chosen delimiter. CSV files are widely supported by various software applications, including spreadsheet software and database management systems. They are commonly used for data exchange and interoperability, making it easy to import and export data between different software tools.
- Spatial TIFF (Tagged Image File Format) is a widely used raster file format specifically designed for storing georeferenced images and spatially distributed data. TIFF files can contain multiple bands or layers of data, such as satellite imagery, aerial photographs, digital elevation models, or other types of geospatial information.
- The netCDF (Network Common Data Form) format is a binary file format used for multidimensional data, in a self-describing and platform-independent manner. It is widely used in the scientific community, especially in the fields of atmospheric and oceanic research, climate modeling, and earth sciences. NetCDF files consist of dimensions, variables, and attributes. Dimensions define the sizes of the data arrays, variables represent the data arrays themselves, and attributes provide additional metadata information about the variables.

Annex II Example of data and level of confidentiality for the Timpaki LL

type	name	variables	owner	use	open	restrictions
Microeconomic inputs	olive oil_pr	Minimum price, Maximum price, Average price	Peza Union (Union of Agricultural Cooperatives of Peza)	PMP, PMAUP	0	Private source
Microeconomic inputs	grape for wine_pr	Minimum price, Maximum price, Average price	Peza Union (Union of Agricultural Cooperatives of Peza)	PMP, PMAUP	0	Private source
Microeconomic inputs	irrigation water_pr	Minimum price, Maximum price, Average price	T.O.E.B. A' ZONE of Mesara (Local Land Improvement Organizations)	PMP, PMAUP	0	Private source
Microeconomic inputs	crops_pr	Minimum price, Maximum price, Average price(price of cucumber, peppers, tomato, eggplant, zucchini)	T.O.E.B. of Ierapetra (Local Land Improvement Organizations)	PMP, PMAUP	0	Private source
	tympaki_basin_wgs	basin	Crete Region	WEAP	0	Private source
Contour lines	elevation_Tympaki_wgs84	elevation lines	Crete Region	WEAP	0	Private source
	Tympaki_area_towns_wgs	towns	Crete Region	WEAP	0	Private source
	Lake_Dam_faneromeni_WFD2000_60_wgs	lake	Special Secretariat for Water	WEAP	1	0
Groundwater bodies	groundwater_wfd2000_wgs	Groundwater	Special Secretariat for Water	WEAP	1	0

Deliverable 6.2 – Data Management Plan

type	name	variables	owner	use	open	restrictions
River	Rivers_WFD_2000_wgs	river	Special Secretariat for Water	WEAP	1	0
River streams location	streams_wgs	streams	Crete Region	WEAP	0	Private source
River Basins	WATER_TYMPAKI1_wgs	River Basins	Crete Region	WEAP	0	Private source
Hydrolithic map	HYDROLITHIC_TYMPAKI_all_wgs	Hydrolithic map	Crete Region	WEAP	0	Private source
Well locations	groundwater_observation_wells_wgs	wells locations groundwater characteristics (CONDUCTIVITY-AVG (μS/cm), WATER TEMPERATURE-AVG (OC), WATER LEVEL, WATER DEPTH (m)), -COORDINATES (lon 24.769, lat 35.058)	Crete Region	WEAP	0	Private source
Groundwater	y_tympaki_aerodromio_daily		Decentralized Administration of Crete OPEKEPE (Payment and control agency for community aid)	WEAP	1	0
Land Use / Land Cover	Crops_final_OPEKEPE_wgs	Crop map	Corine LandCover /	WEAP,PMP, PMAUP	0	Private source
Land Use / Land Cover	CORINE_TIMPAKI_wgs	Land use (Olive trees) Temperature-mean (oC), Temperature-high (oC) , Temperature-high (time), Temperature-low (oC) , Temperature-low (time), Rain (mm), Wind Speed-AVG, Wind Speed-High, Wind Speed-High(time), Wind direction	Corine LandCover /	WEAP,PMP, PMAUP	1	0
Meteorological data			E.M.Y. (Hellenic National Meteorological Service)	WEAP	0	Private source
Meteorological data	PETROKEFALI_Precipitation	Precipitation (mm) - COORDINATES (lon 24.83459, lat 35.03007)	E.M.Y. (Hellenic National Meteorological Service)	WEAP	0	Private source

Deliverable 6.2 – Data Management Plan

type	name	variables	owner	use	open	restrictions
Meteorological data	m_tympaki_a3_monthly	Date,Time,M05 Relative Humidity (AVG),M05 Precipitation (SUM),M05 WIND SPEED (AVG),Duration (SUM),M05 Temperature (AVG),WIND DIRECTION (AVG),Daily ETo,Duration -COORDINATES (lon 24.784, lat 35.084)	Decentralized Administration of Crete	WEAP	1	0
Meteorological data	TEMP_Crete_region	Minimum, Maximum Temperature-Station coordinates	Crete Region	WEAP	0	Private source
Meteorological data	PCP_Crete_region	Precipitation (mm)-Station coordinates	Crete Region	WEAP	0	Private source
River Discharge	koutsoulidis-river	River Discharge (million cubic metres)	Crete Region	WEAP	0	Private source
Faneromeni dam	y_faneromenh_daily	dam characteristics (CONDUCTIVITY (µS/cm), WATER TEMPERATURE (oC), DAM VOLUME (m3), WATER DEPTH, WATER LEVEL) - COORDINATES (lon 24.852, lat 35.094)	Decentralized Administration of Crete	WEAP	1	0
springs	intermittent_springs_wgs	intermittent springs locations	Decentralized Administration of Crete	WEAP	1	Private source
springs	Main_spring_Zaros_Cluster_wgs	Main spring Zaros Cluster locations	Decentralized Administration of Crete	WEAP	1	Private source
Microeconomic inputs	Crops production cost	Average cost of cucumber, peppers, tomato, eggplant, olive, grape & greenhouse	Crete Region	PMP, PMAUP	0	Private source