TransformAr

Accelerating and upscaling transformational adaptation in Europe: demonstration of water-related innovation packages

Data Management Plan

Deliverable 8.3

This project has received funding from the European Union's Horizon H2020 innovation action programme under grant agreement 101036683.

Deliverable Number and Name	D8.3 - Data Management Plan
Work Package	WP8 – Project Management
Dissemination Level	Public
Author(s)	George Sarigiannis (EPSILON)
Primary Contact and Email	Nafsika Papakosta, secretary1@epsilon.gr
Date Due	31/03/2022 (M6)
Date Submitted	29/03/2022
File Name	TransformAr-WP8-D8.3-DMP-v2-29-03-2022.docx
Status	Draft
Reviewed by (if applicable)	Charlotte Francois
Suggested citation	G. Sarigiannis, D. Bakogiannis, M. Bonazountas (2022) Data Management Plan. TransformAR Deliverable 8.3, H2020 grant no. 101036683

This document has been prepared in the framework of the European project TransformAr. This project has received funding from the European Union's Horizon 2020 innovation action programme under grant agreement no. 101036683.

TABLE OF CONTENTS

TABLES

EXECUTIVE SUMMARY

Focus of this Deliverable

The D8.3 – Data Management Plan (DMP) is an initial version of a living document, intended to describe the data management life cycle for all data sets that will be collected, processed or generated by TransformAr.

Deliverable contents

TransformAr deals with many types of data and datasets generated or aggregated thought its various activities. These types of data and datasets will be made available, in line with the guidelines for the Extended Pilot on Open Research Data in Horizon 2020 (European Commission, Directorate-General for Research & Innovation, 2016). The DMP will also include the dataset metadata specification that will be used in the data registry, following an appropriate relevant standard (European Commission, Directorate-General for Research & Innovation, 2016). It will specify the recommended licensing schemes as suggested by H2020. In this first version of the DMP both existing and planned data sets are described.

Conclusions and recommendations

The TransformAr DMP as depicted in D8.3 constitutes the basic tool that will be used to manage data in the TransformAr project.

1.0 Introduction

TransformAr participates in the EU Extended Open Research Data Pilot and as such will deliver a Data Management Plan (D8.3). Open access to research data refers to the right to access and re-use digital research data under the terms and conditions set out in the Grant Agreement. Openly accessible research data can typically be accessed, mined, exploited, reproduced and disseminated under defined and clearly specified terms and conditions (either free of charge or at a cost for the user).

The DMP is not a fixed document; it evolves and gains more precision and substance during the lifespan of the project.

This first version of the DMP is delivered at month 6 of the project. As the project evolves, updates are planned with D8.6 "Data Management Plan first update" at M18 and with D8.7 "Data Management Plan second update" at M36.

2.0 Data summary

2.1 Purpose of the data collected and generated

The overall objective of TransformAr is to demonstrate solutions and pathways, deemed essential for climate and social resilience to achieve rapid and far-reaching Transformational Adaptation (TA). Innovation Packages (IPs), as the combination of Transformational Adaptation Blocks (TABs) and actionable adaptive solutions, are co-developed and demonstrated. Region-Specific Portfolios (RSPs) include Nature-Based Solutions, innovative technologies, financing, insurance and governance models, awareness, and behavioural change solutions.

The project objectives associated with data collection/generation are depicted in the following table:

Table 2.1: Project objectives with related dataset collection/generation

Objectives	Dataset collection/generation
SO1. To demonstrate the potential of co-innovation processes for TA at sub-national demonstrator sites and at EU scale. The demonstration of an innovation ecosystem in demonstrators and a Community of Practice at EU scale aims to create shared ownership of solutions and to increase joint decision-making towards systemic change.	The data is collected through literature review for all the demonstrators and can be used by all the TransformAr partners to understand the characteristics of the selected demonstrators. The data will be retrieved through a desktop review of policy and workshops with key local stakeholders.
SO2. To deliver user-friendly, accessible, and comprehensive multi-sector dynamics data services relevant to transformational adaptation and its water-related challenges and fit to the needs of public and private investors, including citizens in transformational adaptation in the EU Green Deal context.	Weather and climate simulation data will be produced for different scenarios and the data will conclude of hydrological, weather data, hydrodynamic model output, post-processed climate model outputs and environmental data from public repositories. The usable data on the impacts of the transformational adaption on the labour market, activity growth, inter-industrial spillovers will be accessible and will prioritize the EU vulnerable regions, sectors, and communities.
SO3. To develop sub-national transformational pathways that maximize environmental and multi-sector socio-economic benefits and minimizes climate risk losses in view of co-defined quantitative targets on Climate Change Adaptation (CCA).	The data is collected through literature review shedding the light on current and emerging best adaption technologies and practices in KCS addressed by existing practices in Europe and worldwide. The data could be used by TransformAr partners to

	determine adaption solutions. Development of a shared vision and socio-economic pathways of a climate resilient future is presented in the form of a playbook tool and a report presenting adaptation pathways, guidance, methods, and processes considering critical thresholds.
SO4. To test the potential of specific innovations to enable rapid and far-reaching change in the resilience of demonstrators.	In objective S04 there will be generated quantitative tabular data with extensive metadata. Raw and published data (water quality, landcover, crop type, pollution incidents) will be collected. Covered solutions are Nature-Based Solutions, innovative technologies, financing, insurance and governance models, awareness, and behavioral change.
SO5. To accelerate investment for TA across the EU by means of demonstrating bankability and innovative financial schemes.	The data derived from University of Antwerp for WP5 is dependent on the demonstrators and once collected will be processed. The report will cover sustainability profiles, reporting, ex-post assessment of the solutions report and will be reproduced by data mining from other WPs.
SO6. To consolidate a catalogue of solutions and IPs, associated guidance documents, and an understanding of the acceptance and preference of citizens of solutions for transformational adaptation.	The data will be based on the discrete choice experiments which will be conducted in the demonstrator sites. Field experiments will also be conducted, and the raw data will be analyzed, and the project will also transfer the analysis of costs and benefits of adaptive solutions. Ips will be exploited for a large upscaling and feeding of the growing adaptation market.

2.2 Types and formats of data the project will generate/collect

Table 2.2: Datasets collected/generated through TransformAr

WP	Dataset	Type	Format
WP1	Geographic, demographic, economic, climate vulnerabilities, impacts, projections	Qualitative text	Word doc./PDF, SPSS,.xlsx,.csv, JPEG
WP2	Hydrological, weather, environmental, water consumption, biophysical, socio-economic	Quantitative, socio-economic	netcdf, tiff/ESRI, Word doc./PDF, HTML, xml,.mdb,.csv,.tab, .xlsx,.dbf,.shp,.tif, JPEG,.csv,
WP3	Literature review,socio-economic	Socio-economic, Text	SPSS, Word doc./PDF, .html,.csv,.tab, .xlsx, .dbf, .shp, .tif, JPEG
WP4	Numerical, digital image, geospatial data	Quantitative tabular, Qualitative text, Socio-economic, vector & raster, api	SPSS, .xml, .mdb, .csv, .tab, .xlsx, .dbf, .ods, .shp, .tif, .dwg, .rtf, .doc, .html, JPEG, .pdf, json,
WP5	UA & existing data &collected from other WPs	Quantitative tabular, Qualitative text	.csv, .xlsx, .shp, .tif, Word doc., PDF,PPT, JPEG
WP6	Experimental data (9000 observations)	Quantitative tabular, Text	SPSS, .csv, .tab, .xlsx, .rtf, doc., .html, JPEG, PDF
WP7	Retrieve data through review of policy, literature, and workshops Data for dissemination (digital video, image)	Qualitative text	.csv, .xlsx, Word doc., .html, JPEG, PDF, .mp3, .mp4

WP8	Data from UA	Qualitative text	.csv, Word doc., JPEG, PDF, PPT

2.3 Reuse of any existing data and how

TransformAr uses and will reuse existing data where needed throughout the project life. Existing data will be used in various tasks and could be integrated with new data where appropriate. In such cases references to authors and institutions will be made available. Table 2.3 below shows datasets (including literature/journals) used or intended to be used:

Table 2.3: Datasets collected/generated through TransformAr

WP (Task)	Type
WP1 (T1.1)	Literature review and data collected
WP2	Raw data from model/observation / 30 GB generated in first 24 months, public/commercial datasets
WP3 (T3.1)	Literature review (scientific & grey), public/commercial datasets
WP3 (T3.5)	Retrieved from workshops
WP3 (T3.6)	Retrieved from workshops & Literature review
WP4	Raw and published data (water quality, landcover, crop type, pollution incidents) / data generated will not be large / data and information from sensors, written reports (inc. graphs & photos)
WP5	Raw and published data
WP6	Raw and published data / 9000 individual responses for 50 to 100 variables
WP7 (T7.1)	Literature review (technical & grey) and collected data, workshops
WP8	UA data

- What is the origin of the data?
- What is the expected size of the data?
- To whom might it be useful ('data utility')?

Answers to the above questions are integrated in Table 2.4:

Table 2.4: TransformAr origin, expected size, data utility of datasets collected/generated

Dataset	Origin	Expected size	Data utility
WP1(T1.1)	Reuse of existing data	Word doc. 18 MB	Researchers, decision makers, other actors, public
WP2	Weather and climate simulation data Hydrological, environmental	May surpass 1 TB netcdf, tiff/ESRI 20-30 GB	Researchers performing socio-economic analysis, technical personnel supporting policymakers
WP3 (T3.1)	Reuse of existing data	a few MB	TransformAr partners, policy makers, performing organization, public, researchers
WP3 (T3.2)	Site-specific data	a few MB	Decision makers, TransformAr partners, parties working on

			resilience of systems to climate events
WP3 (T3.5)	Data retrieved from workshops	a few MB	Decision makers, local authorities
WP3 (T3.6)	Data retrieved from workshops and literature review	a few MB	Decision makers, governments, parties working on resilience of systems to climate events
WP4	Observational biophysical dataset for the mussel culture, data retrieved from IoT sensors	a few MB up to 7 GB/month	N/A
WP5	Data from UA	a few MB	N/A
WP6	1500 members of the public in each site will participate in these generating choice and public acceptance data	a few MB	N/A
WP7 (T7.1)	Retrieve data through a review of policy, literature, and workshops	a few MB	Decision makers, governments, parties working on resilience of systems to climate events
WP8	Data from UA	a few MB	N/A

3.0 FAIR Data

3.1 Making data findable, including provisions for metadata

The project will be a core contributor to European Open Science Cloud. As such TransformAr will not only provide FAIR data where applicable (data should be "as open as possible and as close as necessary") but will also support the community in the uptake of research data sharing and practices, in alignment with FAIR principles. The following table provides information for making TransformAr data findable, including provisions for metadata to promote reuse.

Table 3.1: Findable data including provisions for metadata

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g., persistent and unique identifiers such as Digital Object Identifiers)?	YES – DOI, unique and persistent URI. Collected services' information will be available at the dedicated portfolio/catalogue system URL. Other aspects of the persistent identifiers could be implemented by periodical snapshots of the database that contains the data.
What naming conventions do you follow?	TransformAr-[WP]-[Deliverable]-[title]-[ver]-[day]-[month]-[year].[ext]
Will search keywords be provided that optimize possibilities for reuse?	YES
Do you provide clear version numbers?	YES – accessed with unique and persistent URI
What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.	Metadata for General Research Data will follow the Dublin Core and DataCite Metadata Schema.

	e.g., For WP3 the following metadata will be provided: title, subtitles, executive summary, graphs (adaptation pathways), figures, Tables, references.

3.2 Making data openly accessible

By default, TransformAr will openly provide data produced following the principle "as open as possible, as closed as necessary", to comply with ethical or security requirements and avoid related conflicting issues. The following table summarizes current considerations.

Table 3.2: Provisions for openly accessible data

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.	Anonymised WP6 survey data and WP1 Stakeholders data. Anonymised WP3 Training Capacity data tool repository, IPR tool repository, TransformAr datasets: A predefined portion of all available metadata will be openly accessible, while authorization will be needed for full access.
How will the data be made accessible (e.g., by deposition in a repository)?	Zenodo for all other datasets. TransformAr data will be made available for human and machine access through the projects' service portfolio/catalogue system.
What methods or software tools are needed to access the data?	WP1 – Survey dataset can be accessed with Microsoft Excel or similar open office tools. Other TransformAr Datasets: The project will deploy and host a service portfolio/catalogue system that will provide full human and machine access to the database that stores information about regional services. Datasets for the software tools will require text editors, pdf viewers, more information will be specified in a following DMP version.
Is documentation about the software needed to access the data included?	NO, unless open-source alternative is not provided – many software choices are available online.
Where will the data and associated metadata, documentation and code be deposited?	Dedicated repository in Zenodo (or else).
Have you explored appropriate arrangements with the identified repository?	Pending
If there are restrictions on use, how will access be provided?	YES – there are GDPR issues associated.
Are there well described conditions for access (i.e., a machine-readable license)?	Conditions for access will be described when the access policies are finalized.
How will the identity of the person accessing the data be ascertained?	Project's AAI integrated with the eduGAIN federation.

3.3 Making data interoperable

Table 3.3: TransformAr provisions for making data interoperable

Are the data produced in the project interoperable, that is allowing data exchange and reuse between researchers, institutions, organisations, countries, etc. (i.e., adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating	Yes, metadata interoperability is ensured by following metadata standards. Widely used standard formats and protocols such as OAI-PMH and .odf will ensure datasets exchange and reusability between researchers.

recombinations with different datasets from different origins)?	
What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?	Standard vocabularies will be used, such as the ASIS&T Thesaurus of Information Science, Technology, & Librarianship. For Stakeholders datasets the standardized vocabulary EU Country Named Authority List (https://data.europa.eu/euodp/en/data/data/aset/country) will be used.
Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?	Yes, although it may evolve dynamically during the project lifetime to ensure an ontology alignment within the EOSC.
In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?	YES, if uncommon or project specific ontologies or vocabularies are used.

3.4 Increase data re-use (through clarifying licences)

Table 3.4: TransformAr data licences and provisions for data re-use.

How will the data be licensed to permit the widest re-use possible?	Probably one of Creative Commons license options, but this will be decided when a full picture of the data will be available. The license for each dataset will be one providing the widest re-use possible. For software, possibly Apache or GNU. More information will be specified in a following DMP version.
When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.	The data will be made available for re-use as soon as the final/publishable version of the data is available.
Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.	This will also be made clear as soon as there is a good view of the datasets achieved.
How long is it intended that the data remains re-usable?	Data is intended to remain re-usable for as long as it is allowed by project resources and infrastructure. This will be reviewed as the project progresses. TransformAr services data will be integrated into the EOSC portfolio system, which will ensure the reuse of collected information.
Are data quality assurance processes described?	Processes are not described, but effort will be put in providing quality data. The following data quality metadata will be provided: accuracy, relevance and consistency for the Stakeholders and Survey datasets.

4.0 Allocation of resources

Table 4.1: Resources allocated for data management and making data FAIR in TransformAr

What are the costs for making data FAIR in your project?	Unknown yet
How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).	This is under consideration and further information will be specified in a following DMP version. Some PMs have been allocated to UA, EPSILON and EQY for this purpose.
Who will be responsible for data management in your project?	Project Coordinator and EPSILON for drafting and updating DMP.
Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?	Resources for long term preservation have not yet been discussed, as this involves the general progress of EOSC and related projects.

5.0 Data security

Table 5.1: Data security provisions in TransformAr

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?	Data collected/generated are stored in a TransformAr cloud workspace such as Teams, where every file stored is maintained and encrypted using AES 256-bit encryption in geographically diverse areas. HTTPS protocol is used for secure communication between endpoints as a standard. It is the usual HTTP which runs on top of encrypted sockets (SSL/TLS) on the transport layer of the network stack (TCP/IP). Data will be moved to Zenodo repositories for long term preservation.
Is the data safely stored in certified repositories for long term preservation and curation?	YES – For the Zenodo repositories all files uploaded to Zenodo are stored in CERN's EOS service in an 18 petabytes disk cluster. Each file copy has two replicas located on different disk servers. For each file they store two independent MD5 checksums. One checksum is stored by Invenio, used to detect changes to files made from outside of Invenio. The other checksum stored by EOS, is used for automatic detection and recovery of file corruption on disks. For TransformAr services data periodical database snapshots will be made and stored independently.

6.0 Ethical aspects

Table 6.1: Ethical aspects related to data sharing in TransformAr

Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).	No ethical or legal issues foreseen for any of the generated datasets.
Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?	YES – TransformAr is requesting compliance for collection and reuse of data. WP6 surveys included a full privacy policy for the protection of natural persons

	concerning the processing of personal data and on the free movement of such data (GDPR). Describes the policies and procedures in place by TransformAr to protect the privacy of users, how the confidentiality of such information is ensured, laws, rights of data subjects and a communication path for further clarifications, should there be needed.

7.0 Other issues

Do you make use of other national/funder/sectorial/ departmental procedures for data management? If yes, which ones?

- NO

8.0 Conclusion

This deliverable is the first version of the TransformAr Data Management Plan and contains an initial description of datasets collected/generated in this first stage of the project. The described datasets may be of value for the project and will be exploited by the different tasks through the course of the project. The document will be updated as the list of datasets is enriched with new information or datasets. Datasets use, sharing, preservation and dissemination aspects will be specified in all cases. All this updated information will be included in the future versions and revisions of the current document.

References

DCC | Disciplinary Metadata. (2021, December 01). Retrieved from DCC: https://www.dcc.ac.uk/guidance/standards/metadata

European Commission, Directorate-General for Research & Innovation. (2016, July 26). *Guidelines on FAIR Data Management in Horizon 2020, v3.0.* Retrieved from OceanBestPractices: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European Commission, Directorate-General for Research & Innovation. (2016, July 26). *Template for the Data Management Plan v3.0.* Retrieved from European Commission | Funding & tender opportunities (SEDIA): https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf#page=10

RDA | Metadata Directory. (2021, December 01). Retrieved from CERIF (Common European Research Information Format): http://rd-alliance.github.io/metadata-directory/standards/cerif.html

Climate change impacts are here and now. The impacts on people, prosperity and planet are already pervasive but unevenly distributed, as stated in the new EU Blueprint strategy (European Commission-EC, 2019). To reduce climate-related risks, the EC and the IPCC agree that transformational adaptation is essential. The TranformAr project aims to develop and demonstrate products and services to launch and accelerate large-scale and disruptive adaptive process for transformational adaptation in vulnerable regions and communities across Europe.

The 6 TransformAr lighthouse demonstrators face a common challenge: water-related risks and impacts of climate change. Based on existing successful initiatives, the project will develop, test and demonstrate solutions and pathways, integrated in Innovation Packages, in 6 territories.

Transformational pathways, including an integrated risk assessment approach are co-developed by means of 9 Transformational Adaptive Blocks. A set of 22 tested actionable adaptive solutions are tested and demonstrated, ranging from nature-based solutions, innovative technologies, financing, insurance and governance models, awareness and behavioral change solutions.

TransformAr

This project has received funding from the European Union's Horizon H2020 innovation action programme under grant agreement 101036683.