



STARS 4 Water

Data Management Plan (2nd release)

Deliverable D7.3

29 March 2024

STARS4Water Data Management Plan (2nd release)

Work Package	WP7 Project Management and Coordination
Due date	M18 (March 2024)
Actual submission date	29 March 2024
Lead beneficiary	Deltares
Lead author(s)	Michiel Blind
Contributors	Gert-Jan Schotmeijer (Deltares), Joost Beckers (VanderSat)

Dissemination Level		
PU	Public	X
SEN	Confidential, only for members of the consortium and the granting authority (including other EU institutions and bodies)	
CI	Classified, as referred to EU Decision 2015/444 and its implementing rules	

Version log			
Version	Date	Released by	Nature of Change
1.0	31-03-2023	H. Duel	Approval final version (first release)
1.7	19-03-2024	M. Blind	(update)
1.8	26-03-2024	Gwyn Reese	(review of updated version)
1.9	26-03-2024	M. Blind	(final updated version)
2.0	27-03-2024	H. Duel	(approval final version)

Citation

Blind, M., Schotmeijer, G-J. & Beckers, J. (2024): Data Management Plan (second release). Horizon Europe project STARS4Water. Deliverable D7.3.

Acknowledgement

This deliverable has been largely inspired by the data management plan of the REACHOUT Project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101036599.



The STARS4Water project has received funding from the European Union's Horizon Europe research and innovation program under the Grant Agreement No 101059372

Disclaimer

The content of this deliverable does not reflect the official opinion of the European Union. Responsibility for the information and views expressed herein lies entirely with the author(s)

Summary

This document is the 2nd release of the Data Management Plan (DMP) of the STARS4Water project. It is an update of deliverable 7.1, released 31-03-2023.

The objective of the Data Management Plan is to describe the data management policies throughout the life cycle of the project, focusing mainly on making data Findable, Accessible, Interoperable and Re-usable (FAIR), but also touching ethical issues related to handling of personal data.

The focus of the data management plan is on output data, which are data originating from within the project. The basic principle of STARS4Water is that all output data generated within the project will be made publicly available according to the FAIR framework facilitated by an adequate application, except data that contains either personal data or strictly confidential information. All software code will be shared if created within the project.

For obvious reasons baseline (input) data will not be shared due to IPR, but metadata from selected open access baseline data will be included in the STARS4Water Metadata Portal. Intermediate data will not be shared.

Personal data will be handled according to General Data Protection Regulation ((EU) 2016/679) and will not be made publicly available.

This document is the second release of the DMP explaining the main concepts and updated with some decisions made since the first release. The document is inspired by the DMP recommendations of the European Commission.

The DMP will continue to be updated under the supervision of the Data Manager (DM) during the project execution as new datasets are defined, existing ones are updated, and the data management policy requires refinement. The DM will also act as Data Protection Officer (DPO).

Another formal update of the DMP is scheduled in M38, D7.4 Data Management Plan (Final Release).

Table of contents

Summary	iv
Table of contents	v
1. Introduction	7
2. Data Summary.....	8
2.1. Data utility and purpose	8
2.2. Data types	8
2.3. Data categorization.....	9
2.4. Metadata.....	9
2.5. Data volume	10
3. FAIR data	11
3.1. Making data findable, including provisions for metadata	11
3.2. Making data accessible	12
3.3. Enhancing interoperability.....	13
3.4. Increasing data re-usability.....	14
3.5. Increased data re-use	14
4. Other research outputs.....	15
4.1. Computer code	15
4.2. Project deliverables and publications.....	15
4.3. Products related to the communication, dissemination and reporting	15
4.4. Non-digitized data.....	16
5. Data management flow and resources.....	17
5.1. Role and responsibilities	17
5.2. Criteria.....	17
5.3. Flow.....	18
5.4. Resources	18
6. Data security	20
6.1. Zenodo, Github	20
6.2. STARS4Water meta-data portal, and STARS4Water Storage environment	20
6.3. Website and e-Learning platform	20
6.4. Other.....	20
7. Personal data and ethics.....	21
8. References	22

Annex 1: Dataset description 23

Annex 2: Output data overviews 27

Annex 3: Data use consent form..... 28

1. Introduction

The STARS4Water project aims to improve the understanding of climate change impacts on water resources availability and the vulnerabilities for ecosystems, society and the economy at river basin scale, including two distinctive elements. First, the need for an international stakeholder community to address their specific information needs and data requirements. Second, the development and application of innovative data services, models, tools to fulfil those needs.

Data are at the core of the STARS4Water project. The focus of the data management plan lies with data originating from within the project. With possibly few exceptions, the data generated within the project will be made publicly available according to the FAIR framework [Ref. 1] facilitated by an adequate application. Existing data will be re-used, new data will be generated and provided to stakeholders. As with all new data generation, this requires careful consideration of Intellectual Property Rights (IPR) and, where appropriate, consideration of personal data protection according to General Data Protection Regulation ((EU) 2016/679) [Ref. 2]. These will be handled following the guidelines set out in this data management plan.

This document is the second release of the DMP explaining the main concepts and updated with some decisions made since the report's first release. The objective of the DMP is to describe the data management policies throughout the life cycle of the STARS4Water project, focusing mainly on making data Findable, Accessible, Interoperable and Re-usable (FAIR), but also touching upon ethical issues related to handling of personal data. This second release of the Stars4Water DMP is updated with some decisions made since the report's first release.

The DMP will be updated regularly under the supervision of the Data Manager (DM) during the project execution, as new datasets are defined, existing ones are updated, and the data management policy requires refinement. A final formal update of the DMP will be in M38 deliverable D7.4, Data Management Plan (Final Release).

2. Data Summary

2.1. Data utility and purpose

The aim of the STARS4Water project is to demonstrate the benefit of existing and new data services to support decision-making in water resources management. STARS4Water will develop and deliver new data services and data driven models for better decision-making in water resources planning at the river basin scale and support actions for adaptive, resilient and sustainable management of freshwater resources in the future. The services will be co-designed with stakeholders to meet their needs, ensuring their relevance and uptake beyond the lifetime of the project. Following a process of co-creation, we aim to provide stakeholders with next generation river basin tools and build a strong Community of Practice.

Data generated by the project will be of interest to researchers in follow-up research, river basin authorities and other groups involved in water resources assessments. Some data may be only created for the geographical scope of the cases, while other datasets will have a broader geographical scope and may therefore be of interest to researchers and water resources experts beyond the case studies of the project.

2.2. Data types

The project will re-use and produce a variety of data of various types, including:

- 1) Geospatial data in various forms:
 - a) Static raster data
 - b) Time-dependent raster data
 - c) Vectorial data
- 2) Time series data

The above “technical” data will be managed according to the approaches described in this data management plan (DMP). The data will be stored, along with descriptive and structural metadata, as described in ‘Annex 1: Dataset description’.

The project will also generate, manage, and collect other artefacts that in a broad sense could also be considered as “data”:

- 3) Computer code, e.g. scripts for data handling, processing or conversion, but also modelling code.
- 4) Formal project deliverables, both public and internal (reports, analysis, scientific publications, minutes, etc.);
- 5) Products related to the project communication, dissemination, and promotion (i.e. photos, videos, presentations, websites, story maps and documents of various formats).
- 6) Non-digitized data (paper records)

Our approach to the management of these types of data will be briefly discussed in Chapter 4, Other research outputs.

The project will handle some personal data, i.e. e-mail addresses of stakeholders, data from potential users that will be gathered through interviews and workshops, registers of users of tools, etc.:

- 7) Personal data: Due to the sensibility of this data, it will be handled internally by each of the beneficiaries and associated consortium partners (together: consortium partners), following the rules set out in national, European Union (EU) and international law about the processing of personal data. This type of data will be addressed in Chapter 7, Personal Data/Ethics.

2.3. Data categorization

The STARS4Water data can be grouped into the following categories:

- 1) Baseline input data includes all data gathered by consortium members working in each of the STARS4Water case studies. This data includes published and open data, for example from Copernicus services and (yet) undisclosed data from (local) authorities and stakeholders. It may be raw, pre-processed and modelled data. Regarding undisclosed data from local (authorities): The project has delivered an internal document on River Basin datasets. These baseline data will be shared internally to facilitate research and development in each case study.
- 2) Intermediate data include data that is analysed and processed in various project activities. This data is used to create the data structures and databases that will underlie the STARS4Water models and tools and the result of e.g. pre-processing of baseline data. Part of this pre-processed data could be of interest for several tools or hubs and hence they will be shared internally where appropriate.
As intermediate data are typically input to models and tools within the scope of STARS4Water they are not considered output data and will not be shared. If preprocessing results in and will not be shared beyond the project. If, for some reason the intermediate data are of added value beyond the project, they may evolve to output data.
- 3) Output data is data that becomes part of the STARS4Water tools that will be shared at various moments, e.g. during dissemination activities, at the end of WP activities, tool releases or at the end of the project. The output data will be relevant to the stakeholders in the River Basin Hubs (case studies) who may want to reuse and refine them. They may be relevant for scientific purposes such as intercomparing / benchmarking of tool and method improvements and for review. Only output data that has long-term value will be made publicly available unless legal restrictions apply.

2.4. Metadata

The output data that is generated within STARS4Water and that has long-term value will be stored along with descriptive and structural metadata.

Descriptive metadata is typically used for discovery and identification, as information to search and locate a dataset, such as title, short description, data source, version, etc. We have collected information for many external datasets in the data inventory (Task 2.1) facilitating finding the data by anyone who is searching for it in databases and catalogues. A template for the descriptive metadata is included in Annex 1: Dataset description.

Structural metadata describes how the components of data are organized. The structural metadata help to interpret the data in the right way and use it in the right manner. This includes information about data type, units, projection, IPR, etc. The STARS4Water project will use standard formats for common data types, which allows using standard visualization and processing tools that are available for these formats. Common data formats that can be used are listed below:

Data type	Example Formats
Static raster	Geotiff, ASCII Raster
Time-dependent raster	NetCDF-CF
Vector	Geojson, zipped shapefiles, KML
Time series (of point measurements)	CSV

For more information on metadata see section 3.3

2.5. Data volume

The total size of the data re-used or generated by the project cannot be estimated at this stage of the project, as both re-use and generation are dependent on the stakeholder needs, availability of local and global data, feasibility of processing and the attainable accuracy and resolution of outputs.

3. FAIR data

All data under the category of “output data” will comply with the FAIR guiding principles: to be findable, accessible, interoperable and reusable. This data management follows the FAIR principles, as published by GO FAIR [Ref. 1].

The next sections describe the procedures and actions defined to follow FAIR principles for output data.

Since the DMP is an ongoing and evolving activity during the project, not all details have been defined at this point but will be refined during project execution.

3.1. Making data findable, including provisions for metadata

This section describes the measures that are taken for making STARS4Water data findable for both humans and computers.

F1. (Meta)data are assigned a globally unique and persistent identifier

All output data of the project will be identified with a unique persistent identifier (PID), for example a Digital Object Identifier (DOI). While the metadata associated to a PID can change over time, its PID will not change. Which PID format will be selected, and which platform will be used to create the PID will be based on the data at hand. Evidently, the PID will be globally unique and persistent.

F2. Data are described with rich metadata (defined by R1 below)

The basic minimum meta-data we apply to datasets are the mandatory properties of the DataCite Metadata Working Group [Ref. 5], aiming to include recommended and optional metadata. The list of properties is included in ‘Annex 1: Dataset description’.

Data sharing platforms typically have their specific mandatory meta-data fields. Zenodo’s metadata is compliant with DataCite’s Metadata Schema minimum and recommended terms, with a few additional enrichments.

Dataset will be described in English with rich metadata. The dataset description template will be reviewed to fit specific requirements detected during project execution.

F3. Metadata clearly and explicitly include the identifier of the data they describe

As metadata are typically located at a different location than actual data, the association between the two will be made explicit by a metadata file and the dataset should be made explicit by mentioning a dataset’s globally unique and persistent identifier in the metadata. This is enforced by the mandatory first DataCite property “Identifier (with mandatory type sub-property)”.

F4. (Meta)data are registered or indexed in a searchable resource

When data are stored on the Zenodo platform they are immediately indexed using OpenAIRE.

In case data are not published on a platform which is automatically indexed, STARS4Water aims to make the data findable by publishing metadata, e.g. using Fair Data Point [Ref. 6].

The STARS4Water metadata portal will also add to the findability of datasets, as metadata of output data with long-term value and metadata of other relevant datasets are stored in the portal.

3.2. Making data accessible

A1: (Meta)data are retrievable by their identifier using a standardised communication protocol.

Within the project the STARS4Water Metadata Portal [Ref. 12] and the project website [Ref.13] have been developed. Both will provide insight in the (meta-) data that the project will produce. The data-storage platform Zenodo is the platform of choice to store and publish output data (see section 2.6.). For each submission, a persistent digital object identifier (DOI) is minted, which makes the stored items easily citeable. (Source: Wikipedia). Zenodo is free. A potential limitation of Zenodo for STARS4Water is the 50GB file size limit. However, this can be mitigated by splitting files, or, if this is not possible, Zenodo is open to discuss larger file size [Ref. 14]

In the unlikely event that Zenodo is inappropriate, the involved partners will seek for alternative means to publish data.

The data will typically be downloadable via the web using common protocols (HTTP, FTP, SMTP...). In case data are not freely accessible an appropriate mechanism to access the data will be installed, involving e.g. authentication and potentially contacting data owners by email.

All necessary material that supplements each output dataset (e.g. software for parsing the datasets, standards documents, etc.) will be provided by the consortium via similar means. This also contributes to making the data interoperable.

The meta-data field providing the conditions for access is called 'Rights' in Annex 1: Dataset description.

Important: For clarity's sake it is important to reiterate that this section only concerns **project output data with long-term value**. Other data, such as baseline datasets, intermediate datasets and any other datasets that will not be publicly available, can be exchanged using the appropriate means such as a NextCloud data storage/exchange facility that has been made available by Forschungszentrum Jülich GmbH, partner in this project.

A2. Metadata are accessible, even when the data are no longer available.

By using Zenodo, STARS4Water aims to provide access to output data with long-term value long after the projects termination. The risk that such data disappear shortly after termination of the project is minor.

As stated, in the unlikely event that Zenodo is inappropriate, the involved partners will seek for alternative means to publish data. In this case specific attention will be given to maintain access to metadata of such datasets. As a first step we will include the metadata as an annex to the data management plan. The second step is to store metadata in the STARS4Water metadata portal [Ref. 12]. To sustain STARS4Water services beyond the lifetime of the project, the metadata platform will be maintained for at least five years after the end of the project.

3.3. Enhancing interoperability

11. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing. This mainly means that (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

Typically, a meta-data portal such as Zenodo will have implemented a formal language and provided meta-data will automatically be transferred to a formal computer language. STARS4Water will follow such examples in case that output data and meta data will not be stored in such external portals.

12. (Meta)data use vocabularies that follow FAIR principles

As elaborated in F2, the basic minimum metadata we apply to datasets are the mandatory properties of the DataCite Metadata Working Group (2021, Ref. 5)), aiming to include recommended and optional metadata. The list of properties is included in Annex 1: Dataset description. Furthermore, we will use domain specific dictionaries, which will be specified in the meta data description field.

The RDA Metadata Standards Catalog is a collaborative, open directory of metadata standards applicable to research data. It is offered to the international academic community to help address infrastructure challenges [Ref. 15]. Most output data generated by STARS4Water will fall in the category of "Earth Science". This category includes metadata standards such as ISO 19115 and the INSPIRE Metadata Regulation.

Although it is a very unlikely situation, if any of the data sets belong to a discipline for which there is no clear metadata standard defined in RDA an appropriate meta-data scheme will be proposed. In such cases it is our intention to publish the generated ontologies or vocabularies to allow reusing, refining or extending them.

I3. (Meta)data include qualified references to other (meta)data.

The basic metadata of STARS4Water output data will include information on relationships to other (meta-) data, e.g. data required to understand the context of the data. We will focus on the relations between different output data of the project and between where relevant between outcome and input data.

3.4. Increasing data re-usability

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes.

One of the objectives of STARS4Water is to stimulate the (re)use of available datasets by river basin authorities. The ability to re-use data means that it must be clear for the users if the data is useful for the application at hand and if there are limitations to its re-use.

To improve reusability STARS4Water will, whenever appropriate include in the DataCite metadata field “description” information on:

- The purpose of data creation
- Important limitations/assumptions in underlying methods
- Method of data generation, models, parameters
- Any other relevant information required to assess the usefulness of the data for re-use.

On the limitations for use, the project aims to make output data publicly and freely available whenever possible and for a minimum of five years after the project terminates. Licenses and terms of use for publicly available datasets will be published in the meta-data. Typical licence standards that will be considered are:

CC0 [Ref. 8] databases, time series, spatial datasets

CC BY 4.0 [REF.9] plots, figures, report

Some datasets may not be made (freely) available due to restrictions imposed by the baseline data. In these cases, the usage restrictions will be explained in the metadata (DataCite field “rights”).

Where relevant, guidance will be given to obtain further information about terms and conditions for use of the data, for example by including a contact person or organization.

3.5. Increased data re-use

To assess usefulness the metadata of outcome will contain information on the scope of the data, the limitations of the data, method of creation, quality aspects, etc. This should allow potential users to assess the usefulness of the data from a content perspective.

In rare cases that non-public third-party baseline data is used then the consent of the owner to re-use / share the data will be confirmed. A basic consent form for is included in Annex 3.

4. Other research outputs

4.1. Computer code

Computer codes includes scripts for data handling, processing or conversion, but also modelling code. For output code, that is new code developed within the project the FAIR approach will apply.

Upon delivery of this first version of the Data Management Plan no decision had been made on coding and documentation standards.

The project delivers software open source. Within the OpenEarth Github community, a restricted STARS4Water community has been created.

Background software will only be shared by the owner of such software.

4.2. Project deliverables and publications

The project will deliver a range of public and restricted deliverable and scientific papers.

Public deliverables:

Quality assured public reports will be available through the project website and, whenever appropriate, the Horizon Results Platform. Deliverables with obvious added value beyond the project consortium / duration may receive a PID and be uploaded to relevant repositories.,

Restricted deliverables:

Restricted deliverables are deliverables which contain sensitive information. They will not be published, nor will their existence be advertised through websites or other means.

Scientific publications:

Peer-reviewed publications resulting from the project will be published as much as possible in open access journals, enhancing reusability. Usually, a PID is automatically assigned to peer-reviewed articles, enhancing findability.

Conference abstracts and articles will be made available through conference proceedings. Unless the conference regulations stipulate otherwise the publications will also be published on the project website.

4.3. Products related to the communication, dissemination and reporting

The project may develop other digital outputs which are not part of previous categories. Such output is typically not useful for re-use or such data is sensitive (e.g. photo's). If reuse is relevant such products are typically integrated in public deliverables and hence no specific strategy is required.

4.4. Non-digitized data

Some baseline data may not be digitally available. In the rare occasion that the project would digitize third-party baseline data, this does not make the digitized data output data and hence they will not be made accessible: Data will flow back to the owner who is responsible for any further publication.

5. Data management flow and resources

This section only concerns output datasets such as modelling results.

5.1. Role and responsibilities

Each **dataset** produced by the project will be the responsibility of one project partner, that will implement the actions to ensure quality and the FAIR principles. In this way, when a partner generates data from a simulation, from the analysis of other data, etc. and it is considered that these data are of interest as a legacy of the project, said partner will oversee its management, which includes quality assurance according to STARS4Water quality requirements.

The Data Manager (DM) will oversee validating the new generated data descriptions (metadata) before updating the DMP. He will also perform the external quality control, that will focus on the requisites established by this DMP.

The process will be refined during the project. At the outset, the envisaged process is as follows:

- 1) Starting at M18, partners generating data will be invited regularly (e.g. each 6th months) to provide an overview of potential output data.
- 2) The Steering Group will review the list and decide whether data have long-term value and will be made publicly available. Publicly available data could still have restrictions, e.g. data can be available but not for free. If such restrictions apply this will be confirmed by the Steering Group
- 3) The partner who generates the output data prepares the data files and metadata based on appropriate (quality) standards (see section 5.2).
- 4) The data manager performs a quality check on (meta-) data. Steps 3 & 4 are repeated till the requested quality has been reached.
- 5) The partner generating the output data uploads the (meta-)data to Zenodo and informs the representative of the STARS4Water metadata portal [Ref. 12].
- 6) The representative of the STARS4Water metadata portal uploads the metadata to the portal.
- 7) The data manager updates the summary table in the EC portal and in 'Annex 2: Output data overviews' of this data management plan.

5.2. Criteria

The criteria for uptake of new data onto the STARS4Water platforms need still to be finalized. At the time of writing of this updated DMP, criteria for data quality assurance have not been established. The availability of first concrete output data sets will facilitate this process.

Criteria may comprise:

1. Accuracy / errors of output data
2. Completeness / data gaps
3. Consistency and other parameters aimed at evaluating the quality of the data.

Data provided by third parties will not be checked in terms of quality control, but quality criteria will be considering quality issues of the baseline data used.

5.3. Flow

Figure 1 provides a graphical overview of the public and internal repositories that STARS4Water consortium will use to manage output data (“data generated”). After quality control the data will either be uploaded to Zenodo. As stated, in the unlikely event that Zenodo is inappropriate, the involved partners will seek for alternative means to publish data (called STARS4Water Storage Environment in figure 1). The external data which includes also input data required to develop output data will not be managed by the STARS4Water project. Only metadata will be included in the STARS4Water Metadata Portal.

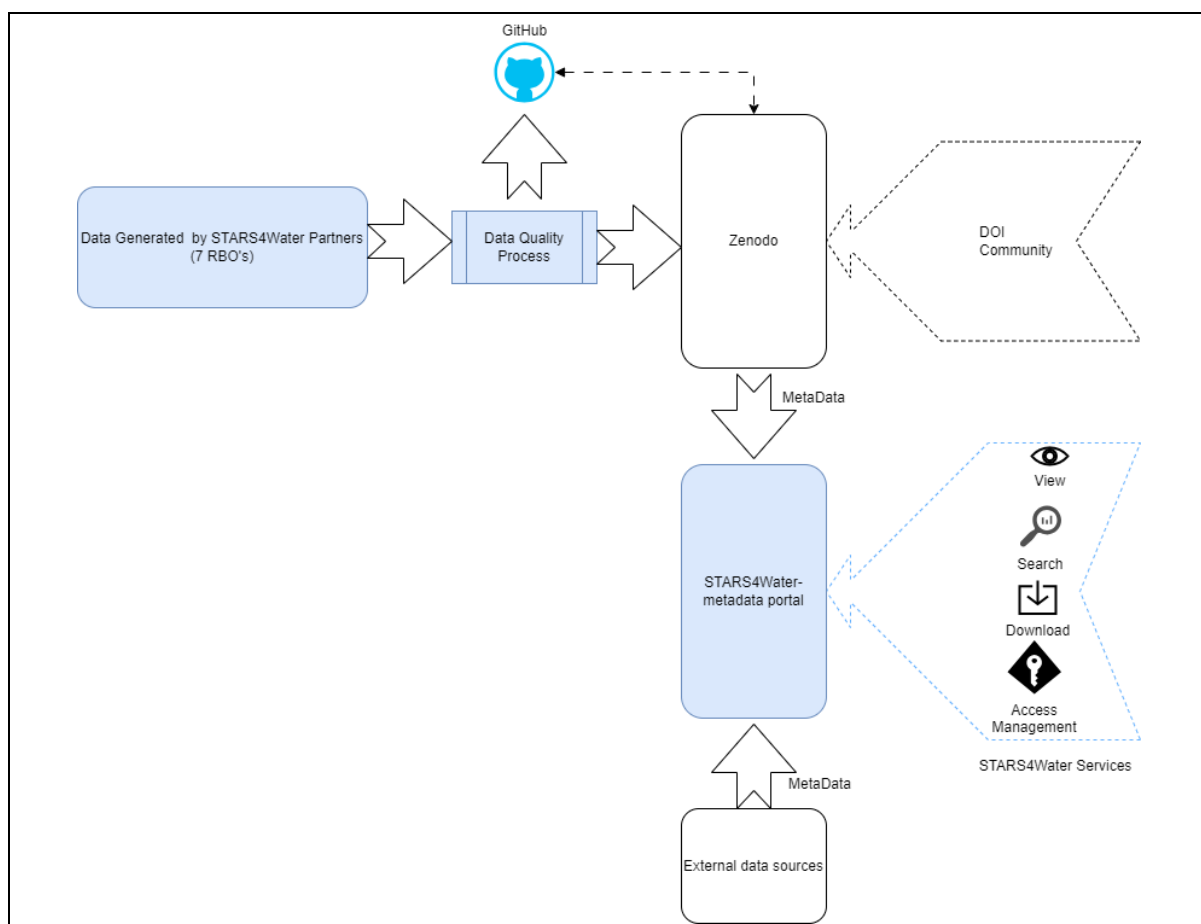


Figure 1. Graphical overview of the workflow of information to support the DMP.

5.4. Resources

The estimated workload and resources required for this workflow is included in the effort and budget assigned to the consortium partners in the Grant Agreement. It is complex to define which part of the activities included in the project are dedicated directly to data related activities as data analysis, software development, etc. and “pure” data management activities (as data upload, etc.).

Due to the use of open repositories, open software, etc. at this stage of the project the estimation of costs associated to licenses, rental of virtual machines, storage space, etc. are estimated to be null. Similarly, the project is not expected to monetize data directly, so the data economy is not expected to generate additional resources.

Concerning open access scientific publications, resources, too, have been included in the project budget. If these resources are depleted consortium partners will seek additional funding.

6. Data security

6.1. Zenodo, Github

STARS4Water uses Zenodo and Github as central data storage and code development platforms. In this case STARS4Water can rely on security and backup systems deployed by these two platforms.

6.2. STARS4Water meta-data portal, and STARS4Water Storage environment

The Metadata Portal is an essential element of the project and project legacy. To some extent this may also be the case for the STARS4Water Storage environment if it is required as addition to Zenodo. STARS4Water will deploy appropriate data security measures ensuring high up-time and backup systems.

6.3. Website and e-Learning platform

The website will contain the public project deliverables. The website hosting organization contracted ensures backup facilities, high up-time levels and prompt handling of security issues and patches.

6.4. Other

To ensure the safety of data and products in terms of recovery, storage and security, in addition to other public repositories and data sharing systems developed within the project each partner will protect all STARS4Water data and products for which it is responsible on premise, utilizing internal data storage facilities, implementing a system of regular backup copies. In this way, data recovery will be possible at any time and all data and products will be stored securely for long-term preservation and preservation. In addition, repositories ensure the secure storage and transfer of sensitive data. This will be described for each dataset in the dataset description template.

7. Personal data and ethics

Within the project we will occasionally collect personal data for example for co-creation workshops, distribution lists, or via non-anonymous questionnaires. This concerns mainly names and addresses and may include opinions. STARS4Water considers such data sensitive. The processing and protection of personal data within the framework of the project will always be subject to the GDPR rules: the consortium partners, including the non-EU consortium partners will comply with General Data Protection Regulation ((EU) 2016/679). In particular, the approach will be that each partner the sensitive/personal data they collect handles according to national laws and European legislation.

Personal data are not output data of the project and hence will not be shared in whatever form.

The STARS4Water project does not plan to collect any personal data that fall into special categories of personal data, such as gender, race or ethnic origin, nor will it collect any data related to religion, political opinion, criminal history or likewise.

As part of collecting user feedback or participating in workshops, interviews, and other outreach activities participation will be voluntary and based on informed consent. This, and other issues regarding personal data, this is further explained in D7.2 Ethical requirements.

8. References

- [Ref. 1] FAIR Principles <https://www.go-fair.org/fair-principles/> (accessed 29-03-2024).
- [Ref. 2] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).
- [Ref. 3] Horizon 2020 Online Manual: Data management https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm#A1-template (accessed 29-03-2024).
- [Ref. 4] Zenodo <https://zenodo.org/> (accessed 15-03-2023).
- [Ref. 5] DataCite Metadata Working Group. (2021). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. Version 4.4. DataCite e.V. <https://doi.org/10.14454/3w3z-sa82>
- [Ref. 6] FAIR Data Point Working Draft, 26 August 2022 <https://specs.fairdatapoint.org/> (accessed 29-03-2024).
- [Ref. 8] Creative Common. CC0 Universal. <https://creativecommons.org/publicdomain/zero/1.0/deed.en> (accessed 29-03-2024).
- [Ref. 9] Creative Commons. Attribution 4.0 International. <https://creativecommons.org/licenses/by/4.0/deed.en> (accessed 29-03-2024).
- [Ref. 10] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm (accessed 26-3-2024)
- [Ref. 11] Stars4Water community on Zenodo: <https://zenodo.org/communities/stars4water/> (accessed 26-3-2024)
- [Ref. 12] STARS4Water metadata portal: <https://stars4water.openearth.nl/geonetwork/srv/dut/catalog.search#/home> (accessed 26-3-2024)
- [Ref. 13] STARS4Water website: <https://stars4water.eu/> (accessed 26-3-2024)
- [Ref. 14] Source: FAQ section <https://help.zenodo.org/> (accessed 15-3-2023)
- [Ref. 15] RDA Metadata Standards Catalog: <https://rdamsc.bath.ac.uk/> (accessed 26-03-2024)

Annex 1: Dataset description

The following table combines essential elements of the DataCite document [Ref. 5], complemented by Zenodo fields and optional inspirational issues which can be addressed in ID17 . The DataCite properties document provides additional details when it comes to additional attributes of the 20 headline properties, including the controlled list's values.

Mandatory fields are highlighted in blue.

ID	Property	Obligation	Explanation
1	Identifier (with mandatory type sub-property)	Mandatory	The Identifier is a unique string that identifies a resource. For software, determine whether the identifier is for a specific version of a piece of software, (per the Force11 Software Citation Principles 11), or for all versions.
1.1	Identifier type	Mandatory	The type of Identifier (e.g. DOI)
2	Creator (with optional given name, family name, name identifier and affiliation sub-properties)	Mandatory	The main researchers involved in producing the data, or the authors of the publication, in priority order. To supply multiple creators, repeat this property.
3	Title (with optional type sub-properties)	Mandatory	A name or title by which a resource is known. May be the title of a dataset or the name of a piece of software (Free text)
4	Publisher	Mandatory	The name of the entity that holds, archives, publishes prints, distributes, releases, issues, or produces the resource. This property will be used to formulate the citation, so consider the prominence of the role. For software, use Publisher for the code repository. If there is an entity other than a code repository, that "holds, archives, publishes, prints, distributes, releases, issues, or produces" the code, use the property Contributor/contributorType/hostingInstitution for the code repository Examples: World Data Center for Climate (WDCC); GeoForschungs Zentrum Potsdam (GFZ); Geological Institute, University of Tokyo, GitHub
5	Publication Year	Mandatory	The year when the data was or will be made publicly available. In the case of resources such as software or dynamic data where there may be multiple releases in one year, include the Date/dateType/ dateInformation property and sub-properties to provide more information about the publication or release date details.
6	Subject (with scheme sub-property)	Recommended	Subject, keyword, classification code, or key phrase describing the resource. Free text.

D7.3 DATA MANAGEMENT PLAN (2ND RELEASE)

ID	Property	Obligation	Explanation
7	Contributor (with optional given name, family name, name identifier, and affiliation sub-properties)	Recommended	The institution or person responsible for collecting, managing, distributing, or otherwise contributing to the development of the resource. To supply multiple contributors, repeat this property.
8	Date (with type sub-property)	Recommended	Different dates relevant to the work, e.g. Date accepted, Date available, date issued...
8.a	dateType		Type of date. If Date is used, dateType is mandatory. Controlled List Values: Accepted, Available,... Copyrighted,...
9	Language	Optional	The primary language of the resource
10	ResourceType (with mandatory general type description subproperty)	Mandatory	A description of the resource.
10.a	resourceTypeGeneral		Controlled List Values, see [ref 5]. Illustrative examples: Dataset, Model, Software, Dissertation,....
11	AlternateIdentifier (with type sub-property)	Optional	An identifier other than the primary Identifier applied to the resource being registered. This may be any alphanumeric string which is unique within its domain of issue. May be used for local identifiers. The AlternateIdentifier should be an additional identifier for the same instance of the resource (i.e., same location, same file). Free text.
12	RelatedIdentifier (with type and relation type sub-properties)	Recommended	Identifiers of related resources. These must be globally unique identifiers.
12.a	relatedIdentifierType		Controlled List Values, see [ref 5]. Illustrative examples: DOI, EAN13, ISBN, ISSN
12.b	relationType		Description of the relationship of the resource being registered (A) and the related resource (B). If relatedIdentifierType is used, relationType is mandatory. Controlled List Values, see [ref 5]. Illustrative examples: IsCitedBy, IsSupplementTo, IsPartOf, ISDerivedFrom
13	Size	Optional	Size (e.g., bytes, pages, inches, etc.) or duration (extent), e.g., hours, minutes, days, etc., of a resource
14	Format	Optional	Technical format of the resource. Use file extension or MIME type where possible, e.g., PDF, XML, MPG or application/pdf, text/xml, video/mpeg.
15	Version	Optional	The version number of the resource

D7.3 DATA MANAGEMENT PLAN (2ND RELEASE)

ID	Property	Obligation	Explanation
16	Rights	Optional	Any rights information for this resource. The property may be repeated to record complex rights characteristics. Free Text. Example: Creative Commons Attribution 3.0 Germany License
17	Description (with type sub-property)	Recommended	All additional information that does not fit in any of the other categories. May be used for technical information. Free text. See below for inspiration.
18	GeoLocation (with point, box, place, and polygon sub-properties)	Recommended	Spatial region or named place where the data was gathered or about which the data is focused
19	FundingReference (with name, identifier, and award related subproperties)	Mandatory For EC projects	Information about financial support (funding) for the resource being registered
20	RelatedItem (with identifier, creator, title, publication year, volume, issue, number, page, publisher, edition, and contributor sub-properties)	Optional	Information about a resource related to the one being registered, e.g., a journal or book of which the article or chapter is part.
Zenodo additional/detailed requirements			
	Keyword(s):		Indicate the keywords describing your dataset. If your dataset is intended to be published in Zenodo, please use its keywords. Include here "HE_STARS4Water" It is mandatory to include "European Union (EU)" and "Horizon Europe"
	Communities:		If your data is published or intended to be published in a repository provide the name and location of the repository
	Cite as		Indicate the proposed citation for your data (e.g. "Jorge Paz & Jorge Paz Jimenez (2022). REACHOUT dataset example (version 1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.6307###)
	License		Indicate the license (Creative Commons, etc.). CC-BY-4.0 international is recommended for reports, graphs, figure, etc. CC0 is recommended for databases, datasets, etc. Take into consideration that this waives all your rights.
Inspirational: Fields to consider addressing in ID 17 (when relevant)			
General information			

D7.3 DATA MANAGEMENT PLAN (2ND RELEASE)

ID	Property	Obligation	Explanation
	Status of the data		Indicate the current status of the dataset: (existing, foreseen, being generated, etc.)
Dataset description			
	Time coverage		Available time range of the data
	Time resolution		Include the amount of time between two data for the same location (e.g. timestep for the outcomes of models, revisiting time for earth observation data from polar satellites, etc.)
	Area coverage		Geographic extension of your data. You can use text ("Global", "Europe", etc. area, etc.) or the coordinates of the corners of your data.
	Spatial resolution		Include the horizontal spatial resolution for raster data, the scale for vector data, etc.
	Data update frequency		Indicate if the data are planned to be updated, if new versions are planned, if it is planned that the data will not be available after a certain date, etc. Leave this cell empty if they are not.
	Metadata		Indicate the files including the metadata if any. Indicate if your data are compliant with any standard. Leave this cell empty if metadata are not provided.
	Data processing		Indicate how your data has been generated
Access, archiving and preservation			
	Back-up		Indicate if data are stored in other mean (e.g. in a RAID managed by the partner producing the data, etc.)
	Preservation period		Indicate how long data will be stored (publicly and not publicly).
	Future actions		Indicate if there are plans for updating the dataset, stop sharing it, etc.
	Recommended use		Indicate the fields of application of your data, recommended processing, etc.

Annex 2: Output data overviews

After Month 18 of the project, this annex will contain a list of data sets shared or anticipated to be shared. Once more matured, data will be available in a table with the following fields. Capitalized fields are requested for EC project reporting.

Title	Text
BRIEF DESCRIPTION	Text
PID	ID
TYPE OF PID	Text
DESCRIPTION OF DATASET	Text
OPEN ACCESS DATA	[y/n]
METADATA OPEN ACCESS	[y/n] Is the metadata of deposited data accessible through open access?
Size [MB]	[MB/GB]
Published	[y/n]
REVIEW PROVISIONS	If data is needed to validate conclusions of a scientific publication, describe the provisions whereby you intend to make it available

As of Month 18, no specific output datasets can be reported. In response to a questionnaire directed to river basin research partners and in month 16 revealed that some potential output datasets are in development for example Lisflood baseline modelling results (Drammen basin) and Machine learning groundwater quality model output (Duero).

Annex 3: Data use consent form

This annex includes a template for Informed Consent that may be used for the managing of third-party data during the execution of the STARS4Water project. The template included in is provided as draft that the consortium partners are expected to refine and adjust to suit their needs when obtaining data. Text that should be always modified are identified as bold.

Dear **XXX**,

You have been asked to share data with the consortium in charge of the STARS4Water project of the Horizon Europe Research Framework Program of the European Commission, grant agreement number 101059372, hereafter ‘the project’. This form of informed consent describes the project and the terms of this cooperation.

STARS4Water aims to improve the understanding of climate change impacts on water resources availability and the vulnerabilities for ecosystems, society and the economy at river basin scale, including two distinctive elements: first, the need for an international stakeholder community to address their specific needs and requirements. Second, the development and application of innovative data services, models, tools.

STARS4Water will develop and deliver new data services and data driven models for better decision-making support on planning actions for adaptive, resilient and sustainable management of fresh water resources, which will be co-designed with stakeholders to meet their needs, ensuring their relevance and uptake beyond the lifetime of the project. Following a process of co-creation we aim to capacitate stakeholders with next generation river basin tools and build a strong Community of Practice.

In this process, the project closely collaborates with several stakeholders to facilitate the deployment of valuable tools and products. These activities involve the compilation, storage and management of datasets and information. Next points describe the procedures for handling the data that you will provide for the implementation of the project, that will be treated with the utmost care.

1. Data will be used for scientific purposes, such as writing scientific articles, developing customized tools, etc. in this case all data will be analysed on an aggregate level and be presented in an anonymous way.
2. Data will be stored in an electronically and physically secure environment at [name of organization storing the data and geographical location where data will be stored].
3. All data will be stored for a period of 5 years after the end of the project.
4. Any data you provide will only be used for the purposes of the STARS4Waterproject.
5. The organization you work for will be mentioned as the source of the data.
6. No original data (understood as non-anonymized data) will be shared with any person or organization, public or private, outside of the Project consortium.
7. However, with your written consent, it is possible to share data with parties who are not members of the consortium but who participate in the city hubs. If you wish to share your data with the participants of your River Basin Hib under the terms of this document, please answer “yes” to the last question at the end of this document. If you prefer to arrange the terms of this eventual collaboration in the future, please, answer “no”. In this case, details of

how data will be shared (which data elements, to/from which stakeholders) will be discussed and specified between involved stakeholders, when this issue becomes relevant.

8. You are free to have your data removed from all the databases associated to the STARS4Water project at any time during its implementation. To do this, simply contact either the project coordinator or your direct contact person with your request.
9. **[Please, enter here any additional point about the data management, collection, licenses, limitations, citation, etc.]**