



STARS 4 Water

Data Management Plan (final release)

Deliverable D7.4

September 2026

Data Management Plan (final release)

Work Package	WP7 Project Management and Coordination
Due date	November 2025
Revised due date	September 2026
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
2.2	11-07-2024	H. Duel	Approval final version (2 nd release)
2.5	17-11-2025	M. Blind	1 st version final release
2.6	22-09-2026	M. Blind	1 st full version final release
2.9	25-09-2026	M. Blind	Review comments provided by contributors and Leandro Avila (FZJ) processed
3.0	28-9-2026	H. Duel	Accepted

Citation

Blind, M., Schotmeijer, G-J. & Beckers, J. (2026): Data Management Plan (final release). Horizon Europe project STARS4Water. Deliverable D7.4. Grant Agreement No 101059372

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 third and final release of the Data Management Plan (DMP) of the STARS4Water project. It is an update of the 2nd release, deliverable 7.3.

The objective of the Data Management Plan (DMP) 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 DMP addresses the handling of three types of data:

- **Technical data, produced by the project.** 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. , Data that contains either personal data or strictly confidential information will not be published. All software code will be shared if created within the project.
- **Technical data, produced by third parties.** One of the key deliverables of the STARS4Water project is a Metadata Portal (D2.8), which contains information about climate data that is produced by third parties, such as Copernicus, NASA, JRC, and others. For obvious reasons, the actual baseline (input) data cannot be shared due to IPR, but the metadata from selected open access baseline data is included in the STARS4Water Metadata Portal.
- **Personal data.** Personal data will be handled according to General Data Protection Regulation (GDPR) ((EU) 2016/679) and will not be made publicly available.

The update in this final release of the DMP concerns a further elaboration of the quality assurance process for metadata and addition of an overview of published output data.

As personal data will not be exchanged and all partners are GDPR compliant, data protection lies with the individual partners (beneficiaries and associated partners). Nevertheless, the DM will, also act as the project's Data Protection Officers' contact (DPO-contact) and all partners are requested to report relevant GDPR issues.

As a final version, text explaining what data actions will be done during the project has been rephrased to what has been done. The final version also includes a list of output data generated by the project.

Table of contents

Summary.....	iii
Table of contents	iv
1 Introduction	1
2 Data Summary.....	2
2.1 Role of data within STARS4Water	2
2.2 Data types.....	2
2.3 Data categorization	3
2.4 Metadata	3
2.5 Data volume	4
3 FAIR data	5
3.1 Making data findable, including provisions for metadata.....	5
3.2 Making data accessible.....	6
3.3 Enhancing interoperability	6
3.4 Increasing data re-usability	8
3.5 Increased data re-use	9
4 Other research outputs.....	10
4.1 Computer code	10
4.2 Project deliverables and publications	10
4.3 Products related to the communication, dissemination and reporting.....	10
4.4 Non-digitized data	11
5 Data management flow and resources.....	12
5.1 Planned and realized process, role and responsibilities	12
5.2 Criteria for including output data in Zenodo.....	13
5.3 Quality Assurance & Quality Control.....	13
5.3.1 Metadata.....	13
5.3.2 Output data	14
5.4 (Meta)data flow.....	14
5.5 Resources.....	15
6 Data security	16
6.1 Zenodo, Github.....	16
6.2 STARS4Water Metadata Portal, and STARS4Water Storage environment	16
6.3 Website and e-Learning platform.....	16

6.4	Other.....	16
7	Personal data and ethics.....	17
	References	18
	Annex 1: Dataset description.....	19
	Annex 2: STARS4Water output datasets.....	23
	Annex 3: Data use consent form.....	25

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 (Figure 1). The focus of the data management plan lies with data originating from project activities. All data generated within the project that is considered relevant for future studies will be made publicly available according to the FAIR framework [Ref. 1] facilitated by hosting on Zenodo (Ref. 2) hosting and including their metadata in the STARS4Water Metadata Portal. Existing and newly generated data is also 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. 3]. These will be handled following the guidelines set out in this data management plan.

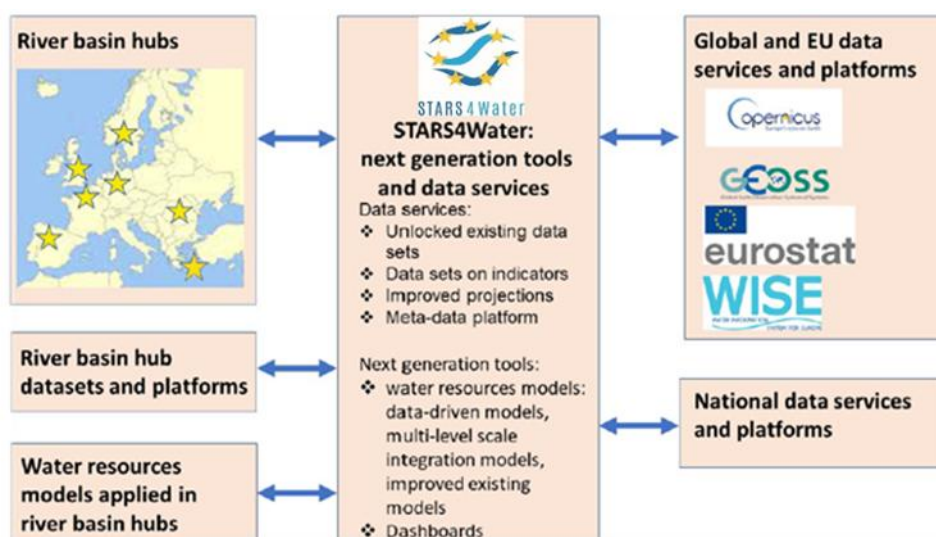


Figure 1. The role of data within the STARS4Water project.

This document is the third and final release of the Data Management Plan (DMP) explaining the main concepts and updated with some decisions made since the report's first and second 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.

2 Data Summary

2.1 Role of data within STARS4Water

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 has developed and delivered 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 were 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 were able 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 was only created for the geographical scope of the use 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 has re-used and produced 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 have been and will continue to 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 also generated and collected 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 has also handled some personal data, i.e. e-mail addresses of stakeholders, data from potential users that were gathered through interviews and workshops, registers of users of tools, etc.:

- 7) Personal data: Due to the sensibility of this data, it was and 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 was 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 it was 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 externally. All preprocessing and intermediate results were 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 is 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 is 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 is 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 uses 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 is difficult to estimate. A total of 490 datasets were included in the Metadata Portal. Around 30 newly generated datasets were uploaded to Zenodo (see Annex 2).

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.

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 are 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. 4], 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.

Each Dataset is described in English with rich metadata. The dataset description template was 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 is 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 third-party data are not published on a platform that is automatically indexed, STARS4Water has made an effort to make the data findable by publishing metadata, e.g. using Fair Data Point [Ref. 5].

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. 9] and the project website [Ref.13] have been developed. Both provide insight in the (meta-) data that the project has produced. The data-storage platform Zenodo has been 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.

A limitation of Zenodo for STARS4Water is the 50GB file size limit. In practice, this has not been a blocking limitation for STARS4Water and no alternative solutions for publishing data were needed.

All data is downloadable via the web using common protocols (HTTP, FTP, SMTP...). All necessary material that supplements each output dataset (e.g. software for parsing the datasets, standards documents, etc.) is 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.

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

By using Zenodo, STARS4Water provides access to output data with long-term value, i.e. even after the project period, thereby minimizing the risk that such data disappear shortly after termination of the project.

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

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

In practice, data usually needs to be integrated with other data and interoperate with applications or workflows for analysis, storage, and processing. This implies that (Meta)data must use a formal, accessible, shared, and broadly applicable language for knowledge representation.

Typically, a data portal such as Zenodo will have implemented such a formal language and provided meta-data will automatically be transferred to a formal computer language. STARS4Water has followed such examples for all output data and meta data.

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. 4)], aiming to include recommended and optional metadata. The list of properties is included in Annex 1: Dataset description. Furthermore, STARS4Water aimed to use domain specific dictionaries.

The RDA Metadata Standards Catalogue 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. 10]. Most output data generated by STARS4Water fall in the category of “Earth Science”. This category includes metadata standards such as ISO 19115 and the INSPIRE Metadata Implementing Rules, which are the result of implementing Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). The INSPIRE Implementing Rules are a profile of the ISO 19915:2003. Other profiles of ISO 19115 in use in European Member States have been made compliant with INSPIRE. Where appropriate, adopting INSPIRE Metadata Implementing Rules is recommended.

Table 1 provides insight into the relation between INSPIRE principles and STARS4Water.

Table 1: STARS4Water and the INSPIRE Principles [16]

INSPIRE common principles	STARS4Water
Data should be collected only once and kept where it can be maintained most effectively.	STARS4Water output data are kept in Zenodo. Updated versions of output data will generally lead to new entries in Zenodo.
It should be possible to combine seamless spatial information from different sources across Europe and share it with many users and applications.	Output data include relevant metadata to facilitate combining data seamlessly. STARS4Water output data does not include APIs to simplify importing data into other processing and visualisation tools.
It should be possible for information collected at one level/scale to be shared with all levels/scales; detailed for thorough investigations, general for strategic purposes.	
Geographic information needed for good governance at all levels should be readily and transparently available.	STARS4Water output data comprise (geographic) information that may be useful for governance. The associated metadata is transparent about the (process leading to the) data.
It should be easy to find what geographic information is available, how it can be used to meet a particular need, and under which conditions it can be acquired and used.	The STARS4Water metadata portal aims at achieving exactly this overview of available information, referencing to both external datasets and to STARS4Water output data.

It has been the STARS4Water policy that, 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 would be proposed. However, this situation has not occurred during the execution of the project.

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

The basic metadata of STARS4Water output data includes information on relationships to other (meta)data, e.g. data required to understand the context of the data. The focus has been 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 has been 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 has, whenever appropriate, included 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.
- Terms of use

Regarding the terms of use, the project has aimed 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 from third parties have been published in the meta-data portal. Typical terms of use for public data are:

CC0 [Ref. 6] databases, time series, spatial datasets

CC BY 4.0 [Ref. 7] 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 are explained in the metadata (DataCite field “rights”).

Where relevant, guidance is 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 contains 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, the consent of the owner to re-use / share the data was confirmed. A basic consent form is included in Annex 3.

4 Other research outputs

4.1 Computer code

Computer codes include scripts for data handling, processing or conversion, but also modelling code. For any new code that is developed within the project, the FAIR approach applies.

The project delivers software open source. Within the OpenEarth Github community, a restricted STARS4Water community [Ref. 8] has been created. Matured software will be published on Zenodo with sufficient data to adhere to FAIR principles and software, like data, can be referenced to using a persistent identifier.

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

4.2 Project deliverables and publications

The project has delivered a range of public and restricted deliverable and scientific papers. The data management policy for these outputs is as follows:

Public deliverables. Quality assured public reports are available through the project website and, whenever appropriate, the Horizon Results Platform. Deliverables with obvious added value beyond the project consortium / duration have received a PID and were uploaded to relevant repositories.

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

Scientific publications. Peer-reviewed publications resulting from the project are 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 were made available through conference proceedings. Unless the conference regulations stipulated otherwise, the publications were also be published on the project website.

4.3 Products related to the communication, dissemination and reporting

The project has produced other digital outputs, which are not part of the previous categories. Such output is typically not re-used, nor is such data sensitive). These products are sometimes integrated in public deliverables. No specific data handling strategy is required.

In case of photo's during events, participants provide consent upon registration or arrival.

4.4 Non-digitized data

Some baseline data may not be digitally available. For example, the project could digitize third-party baseline data. However, this does not make the digitized data a formal project output data and hence they would not be made accessible: These data will flow back to the owner who is responsible for any further publication.

This situation has not occurred during project execution.

5 Data management flow and resources

This section only concerns output datasets such as modelling results.

5.1 Planned and realized process, role and responsibilities

Each **dataset** produced by the project is the responsibility of one project partner, that implements 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.

The Data Manager (DM) will oversee validating the new generated data descriptions (metadata) before updating the DMP.

The process has been refined during the project. At the outset, the envisaged process was as follows:

- 1) Starting at M18, partners (beneficiaries and associated partners) generating data are invited regularly (e.g. each 6th months) to provide an overview of potential output data.
- 2) The Steering Group reviews the list of output data and decides 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 is 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. 9].
- 6) The partner who generated the data, or individual to be appointed uploads the metadata to the portal.
- 7) The data manager updates the summary table in the EC portal and in 'Annex 2: STARS4Water output data' of this data management plan.

Collecting information on expected output datasets proved to be challenging. Several datasets were uploaded to a data portal (like Zenodo) prior to reviewing metadata.

In practice, the following process was followed:

- 1) The partner who generated output data prepared the data files and metadata and submitted the metadata-file to the DM and, at the same time uploaded the data to Zenodo or alternative persistent repository.
- 2) The DM carried out a quality control on the metadata file, providing feedback and/or corrected metadata. (see section 5.3).
- 3) The DM checked if files could be downloaded and opened. In case of many files this was performed based on a sample of files.

This process has proven to be both practical and efficient.

5.2 Criteria for including output data in Zenodo

The process description states: *The Steering Group will review the list (of output data) and decide whether data have long-term value and will be made publicly available.*

Data which will be uploaded in any case are:

- Output data published on dashboards
- Output data used for peer-reviewed articles

Other output data, e.g. data generated by models but not used in publications or portals will only be published if they have long term value.

5.3 Quality Assurance & Quality Control

5.3.1 Metadata

This data management plan focusses making metadata available following the FAIR principles. When applying FAIR principles 'R1. (Meta)data are richly described with a plurality of accurate and relevant attributes' (see chapter 3.4).

Zenodo and other persistent repositories require a (minimum) number of metadata to be provided before for any entry can be published. This information can be quite limited. More extensive data descriptions are typically presented in a 'readme' file or in an accompanying publication.

Annex 1 presents a list of metadata fields clearly marking mandatory fields as such. Other fields in the annex are not mandatory but strongly recommended. To enhance findability of the STARS4Water output data through the STARS4Water metadata portal, the geographic extent of the data was declared mandatory, even though the metadata portal allows for non-geo-referenced data.

The quality assurance of the metadata focused on three aspects:

QA question	Purpose
Are mandatory metadata available?	Ensuring compliance with FAIR principles and needs of different repositories (e.g. Zenodo, STARS4Water metadata portal, EC-reporting portal)
Is the data in the metadata file consistent with the information in the repository entry?	This is to avoid that there are inconsistencies between the repository, the upload to the metadata portal and the EC reporting system. As an example: In some occasions, the title of a resource changed from when it was perceived and entered in the metadata file to when it was indeed published in Zenodo.
Are all mandatory fields filled?	This is to avoid that key fields are missing. As an example, the reference to the funding organisation or the link to the STARS4Water community on Zenodo [Ref. 8] has been omitted several times.

Are the metadata credible/understandable/correctly georeferenced?	This is to check if what was entered as metadata 'makes sense', e.g. names of data producers are linked to the correct organisations. Does the georeferencing make sense given the topic of the data?
Can the data be accessed?	Can files be downloaded and opened (random check in case of multiple files).
Are links correct?	This is to ensure that e.g. the metadata portal refers to the right page in a repository.

It should be noted that many entries in Zenodo also include a read-me file explaining (the structure of) the dataset in more detail or refer to other publications.

5.3.2 Output data

STARS4Water output data are research results. This research is documented in deliverables and/or in peer-reviews publications. Such publications provide the space for extensive discussions on data and modelling uncertainty and propagation of uncertainties during modelling and other forms of processing. Other than the quality assurance of deliverables and peer review of articles, no formal quality assurance assessing aspects such as data accuracy, completeness, consistency, etc, could be carried out within the scope of the project. In other words, the quality and accuracy of output data have not been an element of the project's quality assurance.

5.4 (Meta)data flow

Figure 2 provides a graphical overview of process of publishing output data and retrieving output and other data. The left-hand side visualizes the steps of STARS4Water output data till registering in the metadata catalogue, accessible through the metadata portal. The right-hand side shows that also external data populated the catalogue, by harvesting their published metadata (STARS4Water work package 1 for local data, and work package 2 for European and global data).

The catalogue can be searched, and relevant data can subsequently be obtained from the data publishers.

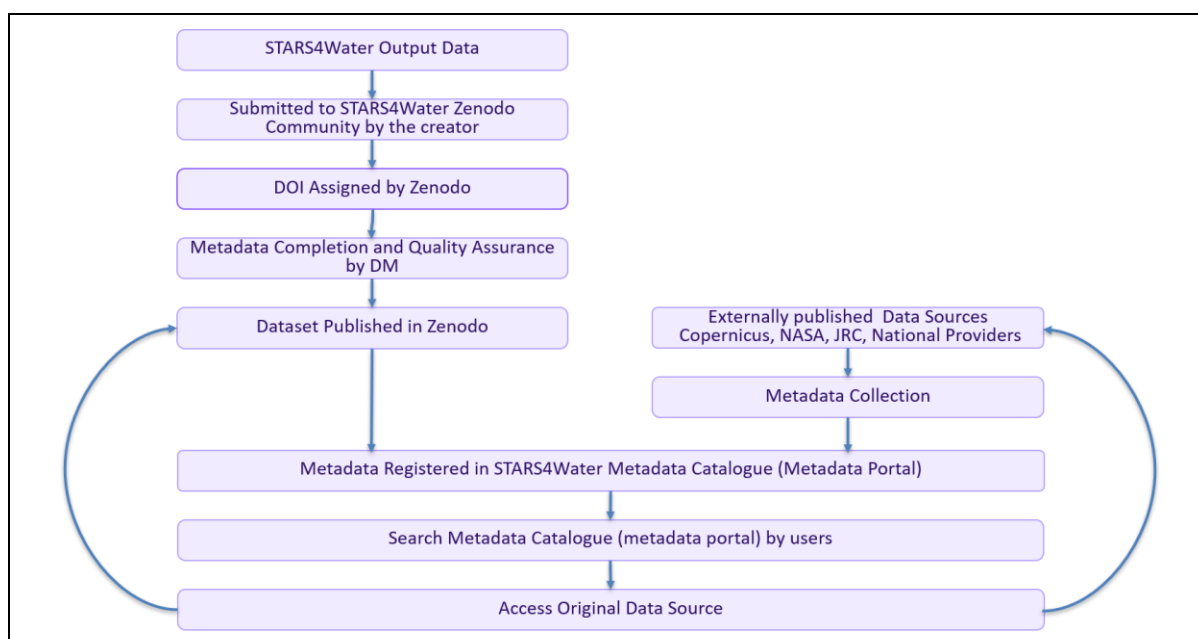


Figure 2. Graphical overview of the workflow of information to support the DMP.

5.5 Resources

The estimated workload and resources required for this workflow was included in the effort and budget assigned to the consortium partners in the Grant Agreement. It is difficult to specify which project activities are dedicated to data processing, such as data analysis, software development, etc. and which activities are “pure” data management (such as data upload, etc.).

In earlier versions of this data management plan, we assumed that there would be no cost associated with licenses, rental of virtual machines, storage space, etc., as we intended to make use of open repositories, open software, etc. This assumption was correct: no additional expenses were incurred for data management. Similarly, the expectation that the project would not monetize on data directly, has been met: no additional resources were generated by monetizing output data.

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. In the final months of the project, Zenodo was “experiencing slowness and intermittent outages due to heavy automated traffic from bots and AI crawlers”. For STARS4water this meant that occasionally work was interrupted.

6.2 STARS4Water Metadata Portal, and STARS4Water Storage environment

The Metadata Portal is an essential tool within the project and project deliverable. STARS4Water partner Deltares has offered to maintain and operate this Portal for the next five years. During this time, Deltares will deploy appropriate data security measures ensuring high up-time and backup systems. After all, the metadata portal is now a component of the Deltares digital infrastructure.

6.3 Website and e-Learning platform

The STARS4Water website contains the public project deliverables and e-Learning recordings. The contracted website hosting organization ensures backup facilities, high up-time levels and prompt handling of security issues and patches. The website will also be maintained for five years after the project’s completion.

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 has protected and will keep protecting 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 is possible at any time and all data and products are stored securely for long-term preservation and preservation. In addition, repositories ensure the secure storage and transfer of sensitive data. During the lifetime of the project, partners have been surveyed twice about any security breaches. No breaches were reported.

7 Personal data and ethics

Within the project, we have occasionally collected 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 has been subject to the GDPR rules: the consortium partners (beneficiaries and associated partners, EU and non-EU) are bound to comply with General Data Protection Regulation ((EU) 2016/679) or in case of partners from non-EU members states (UK and NO) the local equivalent. Committing to GDPR implies that the partner in charge will store the data consistent with GDPR and act upon breaches / data leaks in line with the GDPR. The approach to store data with partner institutes ensures clarity about responsibilities and proper procedures are in place. The Data Protection Officer (DPO) of a partner is responsible for data protection.

In the Deltares' project management team, a data manager has been appointed. The data manager also acts as primary contact on protection of personal data.

Partners are bound to report data breaches to their supervisory authority. In addition are requested to report relevant data breaches directly the coordinator's primary contact on data protection, inter alia the data manager. Partners have been asked twice to confirm that no data breaches have taken place, which they confirmed.

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

In addition, the STARS4Water project has not collected any personal data that fall into special categories of personal data, such as gender, race or ethnic origin, nor has it collected 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 was voluntary and based on informed consent. This, and other issues regarding personal data, is further explained in D7.2 Ethical requirements.

References

- [Ref. 1] FAIR Principles <https://www.go-fair.org/fair-principles/> (accessed 29-03-2024).
- [Ref. 2] Zenodo <https://zenodo.org/> (accessed 15-03-2023).
- [Ref. 3] 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. 4] 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. 5] FAIR Data Point Working Draft, 26 August 2022 <https://specs.fairdatapoint.org/> (accessed 29-03-2024).
- [Ref. 6] Creative Common. CC0 Universal.
<https://creativecommons.org/publicdomain/zero/1.0/deed.en> (accessed 29-03-2024).
- [Ref. 7] Creative Commons. Attribution 4.0 International.
<https://creativecommons.org/licenses/by/4.0/deed.en> (accessed 29-03-2024).
- [Ref. 8] Stars4Water community on Zenodo: <https://zenodo.org/communities/stars4water/> (accessed 26-3-2024)
- [Ref. 9] STARS4Water metadata portal: stars4water.openearth.eu (accessed 25-09-2026)
- [Ref. 10] 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. 4], 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 orange. Items labelled with * are a property of the ISO19115 metadata requirements which is supported by the STARS4Water Metadata portal. The column obligation contains mandatory columns. For STARS4Water Metadata Portal the abbreviation S4W-M or S4W-C (resp. Mandatory and Conditional for S4W Metadata Portal) will be used to indicate the type of obligation.

	Property	Obligation	Explanation
1	Identifier (with mandatory type sub-property) *	Mandatory S4W-C	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 S4W-M	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 S4W-M	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 S4W-M	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 S4W-M	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

D7.4 DATA MANAGEMENT PLAN (FINAL RELEASE)

	Property	Obligation	Explanation
			information about the publication or release date details.
6	Subject (with scheme sub-property) *	Recommended S4W-C	Subject, keyword, classification code, or key phrase describing the resource. Free text.
7	Contributor (with optional given name, family name, name identifier, and affiliation sub-properties) *	Recommended S4W-M	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 S4W-M	Different dates relevant to the work, e.g. Date accepted, Date available, date issued...
8.a	dateType *	S4W-M	Type of date. If Date is used, dateType is mandatory. Controlled List Values: Accepted, Available, ... Copyrighted, ...
9	Language *	Optional S4W-M	The primary language of the resource
10	ResourceType (with mandatory general type description subproperty) *	Mandatory S4W-M	A description of the resource.
10.a	resourceTypeGeneral *	S4W-M	Controlled List Values, see [ref 5]. Illustrative examples: Dataset, Model, Software, Dissertation,
11	AlternateIdentifier (with type sub-property) *	Optional S4W-C	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 S4W-C	Identifiers of related resources. These must be globally unique identifiers.
12.a	relatedIdentifierType	Optional	Controlled List Values, see [ref 5]. Illustrative examples: DOI, EAN13, ISBN, ISSN
12.b	relationType	Optional	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

D7.4 DATA MANAGEMENT PLAN (FINAL RELEASE)

	Property	Obligation	Explanation
14	Format *	Optional S4W-C	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
16	Rights *	Optional S4W-M	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)	Strongly Recommended	All additional information that does not fit in any of the other categories. May be used for technical information. Free text. Strongly recommend including e.g. information on data quality, error propagation.
18a	GeoLocation (with point, box, place, and polygon sub-properties) *	Recommended S4W-M	Spatial region or named place where the data was gathered or about which the data is focused
18b	Coordinate Reference system *	Recommended S4W-M	Coordinate reference system needs to be specified (see link) as assigned by EPSG
	Contiguous rectangle	Mandatory	Referring to 18 a: Specification used in the metadataportal
	West	Mandatory	Rectangle bounding box West-Longitude
	East	Mandatory	Rectangle bounding box East-Longitude
	North	Mandatory	Rectangle bounding box North Latitude
	South	Mandatory	Rectangle bounding box South Latitude
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): *	S4W-M	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"

D7.4 DATA MANAGEMENT PLAN (FINAL RELEASE)

	Property	Obligation	Explanation
	Communities:	= 'STARS4Water'	If your data is published or intended to be published in a repository provide the name and location of the repository
	Cite as: *	S4W-C	Indicate the proposed citation for your data
	License: *	S4W-C	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			
	Status of the data *	S4W-M	Indicate the current status of the dataset: (existing, foreseen, being generated, etc.)
Dataset description			
	Time coverage *	S4W-C	Available time range of the data
	Time resolution *	S4W-C	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 *	S4W-C	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 *	S4W-C	Include the horizontal spatial resolution for raster data, the scale for vector data, etc.
	Data update frequency *	S4W-M	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	Optional	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	Optional	Indicate how your data has been generated
Access, archiving and preservation			
	Back-up	Optional	Indicate if data are stored in other mean (e.g. in a RAID managed by the partner producing the data, etc.)
	Preservation period	Optional	Indicate how long data will be stored (publicly and not publicly).
	Future actions	Optional	Indicate if there are plans for updating the dataset, stop sharing it, etc.
	Recommended use	Optional	Indicate the fields of application of your data, recommended processing, etc.

Annex 2: STARS4Water output datasets

The below table contains an overview of output datasets. These datasets are publicly available. The datasets are available through the persistent identifiers/ links. All datasets are publicly available, predominantly through Zenodo. The metadata have been reviewed (QA) and uploaded to the metadataportal (MDP).

Title/Description	PID
Discharge simulations for the Seine river catchment: the influence of the calibration strategy for projections with a semi-distributed model	10.5281/zenodo.18709621
Future ecological risk due to river flow alteration for the STARS4Water river basins	10.5281/zenodo.15190625
Ecological risk due to future flow alteration across the 7 European STARS4Water river basins for 56 climate and hydrological model ensemble combinations	10.5281/zenodo.21670346
Historical and future water demand for households and industry for the STARS4Water river basins	10.5281/zenodo.14252776
ConvLSTM-based Daily Snow Water Equivalent (SWE) Predictions in US (Rocky Mountains)	10.5281/zenodo.22238156
Discharge simulations for the Seine river catchment, scenarios to analyse the future of low-flows and reservoir management	10.5281/zenodo.17632107
ConvLSTM-based Daily Water Table Depth (WTD) Predictions over Europe	10.5281/zenodo.21406014
Future ecological risk due to river flow alteration in East Anglia, UK using CS-NOW projected river flow data as part of STARS4Water	10.5281/zenodo.21646744
ConvLSTM-based Daily Snow Water Equivalent (SWE) Predictions in the Norway	10.5281/zenodo.21291951
Groundwater vulnerability to agricultural contamination in the Duero River basin	10.5281/zenodo.21293759
Future ecological risk due to river flow alteration in East Anglia, UK using E-FLAG projected river flow data as part of STARS4Water	10.5281/zenodo.21649392
GWAI_duero	10.5281/zenodo.18299839

D7.4 DATA MANAGEMENT PLAN (FINAL RELEASE)

Gridded dataset for groundwater storage estimation in the Duero using a Deep Learning model	10.5281/zenodo.21642181
ConvLSTM-based Daily Water Table Depth (WTD) Predictions over Central Europe (Parflow DE06 domain)	10.5281/zenodo.20703031
ConvLSTM-based Daily Snow Water Equivalent (SWE) Predictions in the Drammen River Basin	10.5281/zenodo.20702459
Future ecological risk due to river flow alteration in the Drammen river catchment	10.5281/zenodo.22233710
ConvLSTM-based Daily Snow Water Equivalent (SWE) Predictions in the Alps	10.5281/zenodo.21257584
RIBASIM-Simulated Danube River Discharge under Reference and "What-If" Scenarios	10.5281/zenodo.21281935
Historical and future irrigation water demand for the STARS4Water river basins	10.5281/zenodo.14246228
Historical and future land use and land cover data for the STARS4Water river basins	10.5281/zenodo.14246388
STARS4Water precipitation deficit for selected river basins in Europe	10.5281/zenodo.15240491
STARS4Water climate indicators - tier 1	10.5281/zenodo.15599187
LSTM and RF models to estimate monthly water table depth anomalies in the Seine river basin	10.5281/zenodo.20701987
Maps of irrigated area of the Rhine basin developed in the research "Identifying irrigated areas using land surface temperature and hydrological modelling: Application to Rhine basin"	10.4121/66647538-ed17-4dd2-9af8-962ed0c61177.v3
ConvLSTM model to predict daily snow water equivalent in the Drammen river basin	10.5281/zenodo.20702458
Data underlying the research of: Improving realism of high-resolution hydrological modelling with anthropogenic water use: a study on the Rhine basin	10.4121/1a382dde-616e-4894-a4b2-222ec45a8a5d.v1
Gauge-adjusted ERA5 Monthly Precipitation and Temperature data for Messara Basin, Crete, Greece	10.5281/zenodo.17130555
Risk of ecological impacts due to flow alteration in the Seine river catchment, under future climate scenarios and reservoir management strategies	10.5281/zenodo.21997573
ConvLSTM-based Daily Water Table Depth (WTD) Predictions in USA	10.5281/zenodo.22237372

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 freshwater 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.]