Deliverable 6.2

Data Management Plan (DMP) version 1.0

Grant Agreement nº: 956623

Project full title: Inventive forecasting tools for adapting water quality management to a new climate

Project acronym: inventWater

Type of deliverable: Open Research Data Pilot

Planned date of deliverable: 31 August 2021

Actual submission date: 31 August 2021

Target Audience: General Public

Authors: Rafael Marcé (ICRA), Daniel Mercado-Bettín (ICRA)

Contributors: Supervisory Board

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623

Contents

1. Data summary

2. Application of FAIR data principles in inventWater

3. Allocation of resources

4. Data security

5. Legal and ethical aspects

6. Other aspects

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623

1. Data Summary

Long-term climate change, extreme events, and seasonal variations in weather have profound impacts on water quality of rivers, lakes, and reservoirs. This implies a pressing need for tools anticipating the impacts of these environmental changes and enabling effective water management that safeguards the ecosystem goods and services freshwater provides. The increasing availability of new meteorological data products and advances in modelling tools now mean that it is possible for the first time to produce reliable forecasts for lake and river water quality on a regional and global scale, an unexploited potential in the water sector. Water managers will therefore profit from the education of a new generation of interdisciplinary trained professionals able to bridge fields such as data science, climate, hydrology, and freshwater ecology, who at the same time develop the necessary skills to translate knowledge and technical novelties into products useful for managers, policymakers, and the general public.

The primary objective of inventWater is to organize a platform providing cutting edge cross disciplinary education of tomorrow's water experts. The core activity is the training program, focused on the development and real-world application of inventive water quality forecasting tools across a range of timescales, to support fast and reliable decision making as well as long-term adaptation policies. inventWater will bring water quality forecasting to the forefront of the scientific disciplines supporting water managers and policymakers to design measures for adaptation to a new climate. The composition of the network and the multidisciplinary of the PhD projects will make tools developed by inventWater fellows applicable to a wide range of pressing water quality issues, from local adaptation to increases of climatic extreme events, to supporting adaptation strategies to climate change and achievement of UN Sustainable Development Goals.

Throughout the 4 years of its duration, inventWater will generate data from a wide range of activities related to modelling efforts activities across ESR projects. In coherence with the objectives of the project and the Plan for Dissemination and Exploitation of the Results (Deliverable 5.1), inventWater results, data, models, tools, and publications will be disseminated so they can be scrutinized by other researchers, potential future partners, and the wider research and operational communities. Consequently, ensuring the legacy and the appropriate access of the data generated by project activities is a priority for the Supervisory Board.

inventWater contributes to the Open Research Data Pilot in the Horizon 2020 programme. inventWater will ensure that its research data is Findable, Accessible, Interoperable and Reusable (FAIR). However, the sharing of data will be temporarily restricted to allow publication releases and PhD accomplishments, considering the “need to balance between openness and protection of scientific information, commercialization, and intellectual property rights (IPR), privacy

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623

concerns, security as well as data management and preservation questions" as indicated in the FAIR data management in Horizon 2020 guidelines published by the European Commission.

The Data Management Plan's (DPM) purpose is to provide the main elements of the data management policy to be used by the inventWater consortium with respect to its complete data cycle. It thus describes 1) types, sources and formats of data generated/collected, 2) the standards to be applied, 3) the data curation methods and 4) the data sharing policies for re-use. The DMP reflects the exploitation strategy as defined in the Consortium Agreement (CA).

The present document is the first version (1.0) of inventWater DMP containing a summary of the datasets (types, sources, formats, WP, and partners), specific conditions to apply for sharing and re-use. **As a continuously evolving document, this DMP will be updated and refined throughout the project lifespan in the form of updated versions of the document.** As a minimum, the DMP should be updated in the context of the periodic evaluation/assessment of the training network.

Most of the tasks in inventWater ESR projects gravitate around modelling, although some productions of primary data are envisaged in some cases (for instance, ESR8). This implies that most data used in inventWater will be sourced from third parties or previous research efforts, and that most data production will involve modelling results. Due to the diversity of topics in inventWater ESR projects, the temporal and spatial extent of the data generated is going to be very heterogeneous across projects, going from data related to single ecosystems or even laboratory experiments to datasets applying to the whole Earth and expanding centuries into the past and the future. Consequently, inventWater DMP should offer a flexible framework to accommodate all the expected diversity of data.

The following tables are an overview of the datasets that inventWater will produce related to activities in the different Work Packages. The tables corresponding to the 15 ESR projects will be produced between the Welcome Retreat (Month 9 of the project) and the submission of the first revision of the DMP (version 2.0, at mid-term project assessment), to have an open discussion about the data used and produced by the different projects before defining the tables.

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623

WP 1 – Innovative forecasting

Data products related to ESR1 to ESR 6 will be updated in DMP version 2.0.

WP 2 – Forecasting for future-proof management and policymaking

Data products related to ESR7 to ESR 10 will be updated in DMP version 2.0.

WP 3 – Identifying adaptation strategies, trade-offs, and synergies

Data products related to ESR11 to ESR 15 will be updated in DMP version 2.0.

WP 4 – Training and Early-Stage Researchers support

Dataset name	Local structured basic cross-cutting training
Purpose and relation with the objectives	Database containing the training opportunities from the different beneficiary and partner institutions that can be accessible for all ESR, independently of the location of the ESR
Work Package	WP4
Partners	ICRA
Data types	Text and hyperlinks
File formats	.txt, .pdf, .ods
Reuse of existing data	N/A
Data production methods	Manual mining of resources.
Expected size	<50 MB
Utility	Very high, for Internal organization of the action
Reuse potential	Very low, may serve as a useful basis in planning training in similar future projects.
Diffusion principles	This data can be public, although the low potential usability does make it an attractive product for other users.

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623

Dataset name	Complementary online training (e-learning)
Purpose and relation with the objectives	Database containing virtual training opportunities worldwide, useful for the various aspects of the training of the ESRs.
Work Package	WP4
Partners	ICRA
Data types	Text and hyperlinks
File formats	.txt, .pdf, .ods
Reuse of existing data	N/A
Data production methods	Manual mining of resources.
Expected size	<50 MB
Utility	Very high, for internal organization of the project
Reuse potential	Very high, other actions and individuals can benefit from the contents of the database
Diffusion principles	The database will be open, posted on the website in the Dissemination tabs and a dedicated blog post.

WP 5 – Dissemination and Communication

Dataset name	Website traffic and visitor statistics
Purpose and relation with the objectives	To monitor the success of outbound information flow, communication, and dissemination.
Work Package	WP5
Partners	ICRA
Data types	Numeric and text Number of visitors (per day, per subpage, per specific item, etc.), number of clicks, number of downloads (as a whole and per items), number of visited sub(pages) per visitor, bounce rate, time spent surfing the website, number of visitors sent by referring sites and arrived from the project social media pages.
File formats	.txt, .csv
Reuse of existing data	N/A
Data production methods	Automatic monitoring/recording of the website traffic
Expected size	<50 MB
Utility	Needed for efficient project communication and reporting.
Reuse potential	Very low, may help planning communication in similar future projects
Diffusion principles	For internal use of the project and public reporting. Data collection will not record any personal or sensitive information about visitors and will be managed in compliance with GDPR.

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623

Dataset name	Social media traffic and visitor statistics
Purpose and relation with the objectives	To monitor the success of outbound information flow, communication, and dissemination.
Work Package	WP5
Partners	ICRA
Data types	Numeric and text Archived statistical outputs about traffic and use of the social media pages (Twitter, LinkedIn) of the project. Mostly metadata and statistical graphs.
File formats	.txt, .csv, .odt, .pdf
Reuse of existing data	N/A
Data production methods	Saving/downloading metadata from the statistical modules of the social media accounts.
Expected size	<500 MB
Utility	Needed for efficient project communication and reporting.
Reuse potential	Very low, may help planning communication in similar future projects
Diffusion principles	For internal use of the project and public reporting. Data collection will not record any personal or sensitive information about the visitors and will be managed in compliance with GDPR.

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623

Dataset name	Dissemination and communication activity database
Purpose and relation with the objectives	Searchable and transparent documentation of the manifold dissemination activities of the project. A dataset for securing all relevant information on scientific papers, conference attendance / presentations, blog posts, press releases, and any other kind of deliverables and dissemination actions.
Work Package	WP5
Partners	ICRA
Data types	Numeric, text, images, presentation files, video files conference attendance data (name, conference data, presentation type, title, file, pictures, etc.), metadata and electronic copies of scientific papers, presentations, and other output documents.
File formats	txt, .docx, .pdf, .csv., .pptx, .jpg, .png, .avi, .mpeg, or equivalent
Reuse of existing data	N/A
Data production methods	N/A
Expected size	50 GB
Utility	Documentation can be used for effective project management and reporting, also for communication and dissemination planning and management
Reuse potential	Low, metadata may serve as a useful basis in planning communication and dissemination strategies in similar future projects.
Diffusion principles	Public database. Will be posted on the webpage in the dissemination tab.

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623

WP 6 – Training Network Management Training Network Management

Dataset name	Event management database
Purpose and relation with the objectives	To ensure effective and well documented meeting/conference/event organization. A dataset for securing all relevant information on self-organized events.
Work Package	WP6
Partners	ICRA
Data types	Numeric, text, images, presentation files, video files Numbers, names and contact data of the participants, schedules, scientific and administrative programs, minutes, presentation files, documentary material (photo, video).
File formats	.txt, .docx, .pdf, .csv., .pptx, .jpg, .png, .avi, .mpeg, or equivalent
Reuse of existing data	N/A
Data production methods	N/A
Expected size	1 TB
Utility	100 GB
Reuse potential	Low, metadata may serve as a useful basis in planning communication and dissemination strategies in similar future projects.
Diffusion principles	Parts of the data can be made public on the webpage. Data collecting and storage will be managed in compliance with GDPR.

Dataset name	Internal contact list
Purpose and relation with the objectives	Contact information of project participants and external advisory board
Work Package	WP6
Partners	ICRA
Data types	Contact information (email, phone number, etc.)
File formats	.txt, .odt, .ods
Reuse of existing data	N/A
Data production methods	Manual collection
Expected size	<50 MB
Utility	Needed for efficient project management
Reuse potential	None
Diffusion principles	None, data will be managed in compliance with GDPR.

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623

WP 7 – Ethics requirements

No data product will be produced by this WP.

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623

2. Application of FAIR data principles in inventWater

The inventWater project will follow the findable, accessible, interoperable, and reusable principles for the datasets presented above and those to be incorporated in upcoming version updates.

2. 1. Making inventWater data findable, including provisions for metadata

Metadata is data on the research data themselves. It enables other researchers to find data in an online repository and is, as such, essential for the reusability of the dataset. By adding rich and detailed metadata, other researchers can better determine whether the dataset is relevant and useful for their own research. inventWater partners will upload, in a standardized form, metadata (type of data, programming language used, etc.) on the project website or the appropriate repository that will allow everyone to find out what data has been provided during the project lifespan. This metadata will be kept separate from the original raw research data.

The bibliographic metadata will include the following:

- The terms the term “Marie Skłodowska-Curie Actions”.
- The name of the action, acronym, and grant number.
- The publication date, and length of embargo period if applicable
- A persistent identifier

Data produced by inventWater will first be stored either in a database by the data owners (personal computer, or on the institutional secure server) or via a centralized database curated by the Coordinator, depending on the nature and size of the database. After an embargo period of 6 months, the data will then be made public on publicly accessible repositories commonly used by the lake modelling community (e.g., ZENODO data repository, the Environmental Data Initiative), including the ISIMIP repositories. By doing so, the inventWater data will be findable via a Digital Object Identifier (DOI) and/or Creative Common’s license numbers will be used as persistent identifiers on open data repositories. Files and folders at data repositories will be versioned and structured by using a name convention consisting of project name, dataset name and ID. The metadata will be built upon the Ecological Metadata Language (EML) using the ezEML web interface unless other formats are required by the specific repository.

2. 2. Making inventWater data openly accessible

The raw data should be openly accessible 6 months after completion of the publishable dataset and a justification will be provided shall this timeline not be respected. A publishable dataset is defined as a dataset which went through processing and analysis to allow producing a manuscript

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623

under its final publishable form. A longer embargo period for the raw data would be allowed if competitive advantages such as the completion of a PhD thesis apply, in which case an embargo of three years will be upheld. Different embargo periods may also apply if imposed by third parties, for instance embargo periods defined by data produced within the ISIMIP framework.

Restrictions on the use of data, software, and code are documented in the inventWater Grant Agreement and can vary according to institutional and national policies and legislation. In case of restrictions on use, metadata will be still provided, which allows for contacting the data owner. The request will then be up for consideration of the data owner and depending on the data owner's decisions full access to the data may be granted.

To get a DOI, data will be deposited in a public repository under 1) open access to data files and metadata and 2) data files provided over standard protocols such as HTTP and use and reuse of data permitted. To protect the copyright of the project knowledge, Creative Commons license will be used in some cases.

Most inventWater data will be produced in common electronic document/data/image formats (.csv, ASCII, NetCDF, .docx, .pdf, .tex, .jpg, .eps, etc.) that do not require specific software. Software developed by inventWater members (e.g., software tools for processing data) will be deposited in code repositories, such as GitHub. Where possible existing software will also be made accessible, e.g., existing R or python scripts.

To preserve and share internal datasets, inventWater will use:

- Individual researchers' data storage media.
- Partner's individual institutions' secure data repositories.

For Open Access data and publications, inventWater will use:

- ZENODO, Environmental Data Initiative, ISIMIP servers, and GitHub where data will be findable through a DOI. Other repositories may be added to this list in future versions of this document.

2. 3. Making inventWater data interoperable

InventWater plans to use controlled vocabularies to describe the metadata fields and support consistent, accurate, and quick indexing and retrieval of relevant data. Keywords and their synonyms will be used for indexing and subject headings of the data and metadata. As controlled vocabularies change within different disciplines of science, these keywords will be developed during the project's course to increase the interoperability of the project's data and metadata. The controlled vocabulary will be added to future versions of this document as Annex I.

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623

2.4. Increase data re-use (through clarifying licenses)

InventWater datasets will be made openly available for re-use. As pointed out above, for reasons of competitive advantages or third party rules a data embargo may apply. For example, in the case of the completion of a PhD thesis, an embargo of three years will be upheld for the raw data (although publishable versions may have to be made open earlier if related to a publication). InventWater projects working with ISIMIP data may need to comply with embargo periods defined by ISIMIP. For publications and publishable versions of the dataset the maximum embargo period defined by the European Commission is 6 months.

Restriction on use of data, software and code are documented in the inventWater Grant Agreement and may vary according to institutional and national policies and legislation. In case of restrictions on use, metadata will be still provided, which allows for contacting the data owner. The request will then be up for consideration of the data owner and depending on the data owner's decisions full access to the data may be granted.

Creative Common Licensing with be used to protect the ownership of the datasets. Both Share-Alike and NonCommercial-ShareAlike licenses will be considered for the parts of datasets for which the decision of making that part public has been made by the Consortium.

The Supervisory Board will ensure the proper conduct of the project's data management.

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623

3. Allocation of Resources

The costs for making data FAIR will be covered as follows:

- Overall coordination and assessment of the status and fulfilment of the DMP by the Project Manager, included in the Common pot (see Consortium Agreement).
- Fees associated with the publication of scientific articles containing project's research data in "Gold" Open access journals. The cost sharing among beneficiaries, in the case of multiple authors, shall be decided among the authors on a case-by-case basis.
- Project Website operation: covered by the Common pot (see the Consortium Agreement).
- Data archiving at public repositories: free of charge.
- Copyright licensing with Creative Commons: free of charge

The Common pot of InventWater will not cover data management resources other than Project Management, website maintenance including updating of links to the distributed institutional repositories. Costs for supporting people/institutions in making data open access are not covered by this and should be provisioned with the corresponding institutional costs.

Every inventWater partner is responsible for the data they produce. We intend to provide access to the open data indefinitely; however, curation of the inventWater web portal after the project end is not included in the Grant Agreement provisions and will be funding dependent. Therefore, long term storage of data on the website will be discouraged.

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623

4. Data Security

None of the data generated in this project are considered sensitive data, thus data security regulations are deemed not necessary.

InventWater will use methods that promote easy access, extended contact, and trust building among participants. The following guidelines will be followed to ensure the security of the data:

- Store data in at least two separate locations to avoid loss of data.
- Encrypt data if it is deemed necessary by the participating researchers.
- Limit the use of USB flash drives.
- Label files in a systematically structured way to ensure the coherence of the final dataset.
- Use control versioning software whenever appropriate.

Long-term data preservation security will be ensured by beneficiaries' institution data repositories and the open data repositories used during the action.

Personnel data in inventWater will be managed in compliance with GDPR. However, this is limited to internal data for project communication, and no personal data collection for research purposes is envisaged.

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623

5. Legal and ethical aspects

InventWater partners need to comply with the Ethics on research integrity as described in the Description of Action. In addition, inventWater partners have to comply with national or international legislation related to data collection (e.g., Nagoya). These include, e.g., legislation on animal experimentation, legislation on privacy of human trial subjects, and legislation on experimentation on exotic and/or invasive species.

Rules about ownership of the results, joint ownership and intellectual property rights are detailed on the relevant sections of the Grant Agreement and the Consortium agreement. Additionally, further information about the dissemination of results can be found in the Plan for Exploitation and Dissemination of Results (PEDR).

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623

6. Other aspects

The responsible Data Manager of inventWater is Rafael Marcé (ICRA).

HISTORY OF CHANGES		
Version	**Publication date**	**Changes**
1.0	31.08.2021	Initial version

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623

ANNEX I. Controlled vocabulary

To be updated in version 2.0 of this document

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956623