SOS-Water deliverable report

D7.4 Data Management Plan

Lead beneficiary	1- IIASA	Due Date	30 September 2024
WP no	7	New due date (if delay)	
Task no	7.4	Actual Delivery Date	
Dissemination level	PU – Public	Status	Updated

Authors

Authors	Partner no	Partner organisation	Name of author
Main author	1	IIASA	Taher Kahil
Supporting author	1	IIASA	Silvia Artuso

Review

Authors	Partner no	Partner organisation	Name of author
Technical review	2	UU	Jen Steyaert
	9	EUTEMA-RS	Alberto Fresolone

Document history

Date	Version	Chapters affected	Description of change	Author	Document status
19/08/2024	0.1	All	Outline	Silvia Artuso	DRAFT
11/09/2024	0.2	All	First draft	Silvia Artuso	
12/09/2024	0.3	All	First revision	Taher Kahil	DRAFT
17/09/2024	0.4	All	Second draft	Silvia Artuso, Taher Kahil	DRAFT
25/09/2024	1	All	Finalisation & formatting	Silvia Artuso Alberto Fresolone	FINAL
14/01/2026	2	All	Update to include deviation from Open Access standard	Silvia Artuso	UPDATED

Executive summary

This report describes the updated Data Management Plan (DMP) of the SOS-Water project. The purpose of this DMP is to provide an overview of all datasets collected and generated by the project and to define the management strategy of these datasets during and after the end of the project with the purpose of making research data Findable, Accessible, Interoperable and Re-usable (FAIR). The SOS-Water DMP follows the structure of the Horizon Europe DMP template. It reflects the status of the data that is collected, processed or generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved. This version of the DMP updates the general strategy of data management in SOS-Water for handling data management-related issues on the administrative and technical levels and provides more details on topics like data and metadata collection, publication and deposition of open data, and the data repository infrastructure. This DMP is a living document that will continue to be updated as the project implementation progresses and when significant changes occur.

Table of Content

List of Tables and Figures

1. Data Summary

Open science practices are central to the successful completion of SOS-Water. They are embraced by the project coordinator and partners, who already follow open-science and data policies in all their research. All tools, models, data, and methods developed within SOS-Water are being implemented following open-science principles while complying with the obligations for data protection by the GDPR and any possible restrictions imposed by the data providers. As a truly interdisciplinary project, SOS-Water highlights the importance of open source and open science in the context of water system modelling, as creating a water system SOS assessment framework requires models and approaches that are accessible, reproducible, and modifiable.

Throughout the four years of its duration, SOS-Water will generate data from a wide range of activities related to modelling efforts, collection of Earth Observation data, and stakeholders' engagement processes. In coherence with the objectives of the project and the Plan for Dissemination and Exploitation of the Results (Deliverable D6.1), SOS-Water results, data, models, tools, and publications are and will be disseminated so they can be scrutinized by other researchers, potential future partners, and the wider research and operational communities.

This DMP version provides an updated overview of the data that is needed and generated in SOS-Water. This includes the type and format, the purpose, the size, and finally the origin of the data. In addition, the potential of the data to be used outside of the project will be explained. In the following sections, we respond to the questions provided in the Horizon Europe DMP template.

Data re-used or generated and its purpose

SOS-Water is collecting, processing, storing, using and generating a large amount of data related to water resources globally and in different case studies. These data are and will be used to: (i) improve the parametrization of SOS-Water models; (ii) add new functionalities to these models, for instance by increasing spatial and temporal resolutions or improving water use schemes; (iii) facilitate model integration; and (iv) inform SOS-Water indicators. To handle this data, the SOS-Water consortium provides the necessary expertise, facilities, and equipment to develop, process, and host existing and generated data.

To support the objectives outlined above, several datasets have been acquired from open-source databases, as detailed in Table 1 below. These datasets include: (i) widely used climatological, hydrological, socio-economic, and biodiversity data; (ii) climate model outputs and forcing data; and (iii) water-related Earth Observation (EO) data. The availability of state-of-the-art EO datasets for water resource monitoring in SOS-Water was reviewed in Deliverable D3.1, "*Data inventory and EO data needs for water resources monitoring Improved EO application prototypes*" completed in March 2023. These datasets are currently either stored locally by the relevant SOS-Water partners or are available for the entire consortium in the web-based collaborative platform of the project on Microsoft SharePoint. This platform is hosted by the coordinating institute IIASA and each partner has access to it, based on a previous identification with email and password. It is important to mention that many of the SOS-Water models (e.g., PCR-GLOBWB, CWatM, CASCADE, and Hydro-economic model of the Jucar water resources system) have developed their databases over many years, which are being and will continue to be further enhanced within SOS-Water. For usage at the local level, global datasets are then cropped and adapted to each of the case study area, as specified in Deliverable D2.1, "*From Global to Local Scenario Narratives*", with data sources referenced in its Appendix I. Additionally, case study-specific data will be collected from various local sources to improve the models' relevance at the local level (e.g., data from the International Commission from the Protection of the Danube river (ICPDR) for the Danube, from the Centre for Hydrological Studies of Spain and the Jucar River Basin Agency for the Jucar, from the Mekong River Commission for the Mekong). Additionally, surveys for obtaining suspended sediment data in the Mekong Delta are currently being implemented locally by the SOS-Water partner SIWRP. More information about case-study specific data has been provided in

Deliverable D2.2, "*Established linkages between WSMs and IMs and building and benchmarking IWMS*" completed in September 2024.

Table 1 Input data from global, open-access databases used in SOS-Water.

Type of data	Examples of databases
Hydrological data	- River discharge from the Global Runoff Discharge Centre [http://grdc.bafg.de/] - Evapotranspiration accessed with Earth Engine [https://developers.google.com/earth-engine/datasets/catalog/NASA_GLDAS_V021_NOAH_G025_T3H#bands] - Hydrography90m - globally seamless, standardised hydrographic network, with corresponding stream topographic and topological information [https://hydrography.org/] - Global gridded soil information [https://soilgrids.org/] - Access to water-related spatial data for the different case studies: Danube basin [https://www.danubegis.org/], Jucar basin [http://saih.chj.es/], Mekong basin [https://portal.mrcmekong.org/], and Rhine [https://www.govdata.de/]
Biodiversity data	- Global Biodiversity Information Facility (GBIF) [https://www.gbif.org/] - Data housed in fish collections in natural history museums, universities and other institutions (Fishnet2 Data) [http://www.fishnet2.net/] - Integrated Digitized Biocollections (idigbio) [https://www.idigbio.org/] - VertNet [http://www.vertnet.org/]
Socio-economic data	- Population and GDP projection data from the SSP database [https://tntcat.iiasa.ac.at/SspDb] - GDP projections from Murakami et al. (2021) [https://gcp-tsukuba.github.io/SSP-downscale/] - Populations projections from the Socioeconomic Data and Applications Center (sedac) [https://sedac.ciesin.columbia.edu/data/set/popdynamics-1-km-downscaled-pop-base-year-projection-ssp-2000-2100-rev01/data-download] - Reservoir data from the Global Reservoir and Dam Database (GRanD) [https://sedac.ciesin.columbia.edu/data/set/grand-v1-reservoirs-rev01] and the georeferenced global dams and reservoirs dataset (GeoDAR) [https://essd.copernicus.org/articles/14/1869/2022/] - Spatially explicit global database on wastewater treatment plants [https://www.hydrosheds.org/products/hydrowaste] - Future land use from the Land-Use Harmonization 2 (LUH2) project [https://luh.umd.edu/data.shtml] - Crop distribution from the Spatial Production Allocation Model (SPAM) [https://mapspam.info/]
Climate data	- Climate reanalysis data from ERA5, ERA5-Land, UERRA or CERRA [https://cds.climate.copernicus.eu/]. - Historical and SSP-RCPs scenarios temperatures and precipitations from CMIP6 climate projections [https://cds.climate.copernicus.eu/api/v2] - Climate data at high resolution for the earth's land surface areas (CHELSA) [http://chelsa-climate.org/] - Bias-corrected climate-input data provided within the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP), based on CMIP6 climate model data [https://www.isimip.org/protocol/3/]

	- Temperature and precipitations anomalies from the IPCC Sixth Assessment Report [https://interactive-atlas.ipcc.ch/] - Improved regional climate change projections from EURO-CORDEX [https://www.euro-cordex.net/]
Earth Observation data	- Copernicus Global Land Service [https://land.copernicus.eu/] - Several climate variables from ESA CCI Climate Dashboard [https://climate.esa.int/] - ESA Land Cover Data [http://maps.elie.ucl.ac.be/CCI/viewer/download.php] - GRACE satellite products accessed with Earth Engine [https://developers.google.com/earth-engine/datasets/catalog/NASA_GRACE_MASS_GRIDS_LAND] - The Global Gravity-based Groundwater Product (G3P) [https://www.g3p.eu/]

Types and formats of data generated or re-used

SOS-Water is generating a diverse range of output comprising both improved modelling tools and simulated model outputs related to water resources availability and use and water values needed to assess and estimate the range of identified water system indicators and meta-indicators.

SOS-Water adheres to standard data formats used within the environmental research communities, including tabulated (CSV, Excel), raster (GRIB, NetCDF, GeoTIFF) and vector (Shapefiles, GeoJSON) data. To deliver and visualize results, SOS-Water is developing dedicated visualization tools to facilitate the exploration of the SOS by combining existing multi-dimensional visualization techniques (Visual Data Mining, Visual Analytics) with narrative methods for collective sense-making. These tools are designed to support stakeholders and policymakers to conceive, develop, and appraise alternative policies. Such tools will support the definition of storylines and pathways that consider climate and socio-economic conditions, specifying the ones that generate performance failures, thus supporting the translation of complex models into actionable plans. Deliverable D5.3, "*Safe Operating Space visualization tool,*" is expected to be available by the end of September 2025, and a first prototype is being created for Milestone 17 in September 2024 (accessible at this link).

Size of the data generated or re-used

The data size that is being generated or re-used depends, among others, on the data type, spatial resolution, and the number of scenarios and models to be run. The size can range from several MBs for socio-economic statistical data to multiple GBs for daily high-resolution EO products. For example, the dataset, currently stored in Zenodo, containing updated watershed boundaries produced by the paper Burek & Smilovic (2022) is 143.5 MB. Another dataset, also stored in Zenodo, containing the code, input and output data used in the paper Tangi et al. (2022) is 73.5 MB. One of the specific objectives of SOS-Water is to reduce the number of water indicators and introduce new meta-indicators that represent the different dimensions of the water system. Achieving that objective substantially reduces the associated data output and long-term storage requirements from a data management perspective, and a total data volume of less than 5TB is expected.

Data utility outside SOS-Water

Data generated in SOS-Water has a high potential to be used outside the project. Collated input data, model codes and simulation results produced in SOS-Water are expected to be useful for scientists working on other environmental-economic assessments (e.g., agriculture, energy, climate change, sustainable development). Synthesized outputs and visualizations will be useful for water resource

managers, water users, policymakers, and the general public interested in the sustainable use of water resources.

2. Fair Data

2.1 Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

The data generated in SOS-Water is planned to be shared using institutional (e.g., IIASA-PURE, EAWAG-DORA, UU-YODA, UPV-RiuNet) and public (i.e., ZENODO) repositories, which assign persistent identifiers. Model codes and input data of reasonable sizes will be available on GitHub, which identifies each update with a unique code.

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Rich metadata is provided in SOS-Water, providing descriptive information about the shared data (e.g., origin, length, timestep, format, geo-reference, units, etc.). Readme files are used in SOS-Water to provide such information. Moreover, standard data formats for climatic and hydrological data, such as the NetCDF CF (Climate and Forecast metadata conventions) compliant data standard, are used in SOS-Water, which holds georeferenced maps through time, including data description and units. Ecological and biological data will also follow metadata standards as the GBIF Metadata Profile (GMP).

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

We provide search keywords when depositing metadata in the selected institutional and public repositories.

Will metadata be offered in such a way that it can be harvested and indexed?

Metadata provided through the selected institutional and public repositories can be harvested and indexed.

2.2 Making data accessible

Repository:

Will the data be deposited in a trusted repository?

Most models used within SOS-Water are open-source (e.g., CWatM, PCR-GLOBWB 2.0) or are accessible promptly upon request (e.g., Jucar River Basin model). Codes of open-source models are available via hosting platforms such as GitHub to ensure continuous development and version control. For example, the Community Water Model (CWatM) is fully open source, and its source code is freely available on GitHub. SOS-Water pursues to advance water use modelling by developing a new, dynamic water use scheme. The scheme's associated development, documentation, and publication adhere to the open-source principles embraced by the project partners. The modelling tool and development process are made available and provided via GitHub as a supplement to scientific publications detailing and validating the new scheme.

SOS-Water is producing a diverse range of model results and data to estimate the identified indicator sets and meta-indicators – both under current conditions and different future scenarios and management pathways. SOS-Water will obtain (high-resolution, gridded) maps of indicators and

provide those under FAIR data sharing standards to maximize uptake by the wider scientific community and ensure its legacy beyond the project duration. The data is and will be shared using selected institutional and public data repositories while complying with the obligations for data protection by the GDPR and possible restrictions imposed by the data providers. On Zenodo, a SOS-Water Community page has been created where we are collecting all publications, data and metadata produced within the framework of the SOS-Water project. For larger datasets that are difficult to store on Zenodo (e.g., datasets for historical and projected business-as-usual simulations generated by the models that ran in each case study) a dedicated repository has been created on the YODA platform, maintained by Utrecht University, which is OpenAIRE-compliant and freely accessible upon authentication, according to the FAIR principle (point A1.2). The selected data repositories guarantee the long-term availability of its content through a commitment of the hosting partners and institutions. The data is also made available as supplements to scientific publications that document and validate the respective output.

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

Access to the identified institutional and public repositories is provided for free, and they have no size limits. The different partners have already been using them in previous projects.

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

The selected institutional and public repositories provide Digital Object Identifiers (DOI) to identify, share, publish, and cite the data.

Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

Open science practices are central to the successful completion of SOS-Water and are embraced by the project partners, who already follow open-science and data policies in all their research. All tools, models, data, and methods developed within SOS-Water are implemented following open-science principles.

Most models, tools, techniques, and algorithms used in SOS-Water (many of those developed by members of the consortium) are open-source (i.e., their source codes are freely available and can be openly accessed, used, shared, and modified). As a truly interdisciplinary project, SOS-Water highlights the importance of open source and open science in the context of water system modelling, as the creation of a water system SOS assessment framework requires models and approaches that are accessible, reproducible, and modifiable. New water use schemes developed within SOS-Water, and other modelling advancements will therefore follow open-science principles, as the models will be made available to members of the scientific community and others. It is important to mention that data generated in SOS-Water is always published at a high level of aggregation, ensuring data protection.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Data made available through SOS-Water is validated before open access is provided to all users. We will also aim to share intermediate data if an embargo is applied.

Will the data be accessible through a free and standardized access protocol?

The data is and will be freely accessible through the selected institutional and public repositories, which follow standard protocols such as HTTP and OAI-PMH.

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

We do not foresee any restrictions on the use of the data.

How will the identity of the person accessing the data be ascertained?

Data shared through the selected institutional and public repositories are publicly accessible and licensed under the public domain. Therefore, no authorization is required to retrieve it. However, the protocols of these repositories allow for an authentication and authorization procedure, where necessary.

Is there a need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?

At this stage, we do not foresee the need for a data access committee, as data is shared while complying with the obligations for data protection by the GDPR and any possible restrictions imposed by the data providers.

Metadata:

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

Metadata is made openly available, and we do not foresee any issues with the data having a CC0 agreement associated with it. For instance, Zenodo metadata is licensed under CC0. Moreover, we always ensure that metadata contains the required information to enable the user to access the data.

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

We expect the data to remain available and findable for the lifetime of the selected repositories. In the case of the data becoming outdated, which is possible with additional updates to the data/products, previous versions will be noted.

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

Documentation and references about software are provided in case needed to access or read the data. Codes for processing the data and of open-source models used in SOS-Water are and will be made available via GitHub.

2.3 Making data interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

To ensure interoperability, SOS-Water adheres to standard data formats used within the respective communities, that includes using tabulated (CSV, Excel), raster (GRIB, NetCDF, GeoTIFF) and vector (Shapefiles, GeoJSON) data. For example, climate and hydrological data is provided using NetCDF, which consists of self-describing, machine-independent data formats that facilitate the exchange and reuse of scientific data. Many (open source) software applications do exist that are able to read and generate NetCDF datasets.

In order to deliver and visualize results, SOS-Water is developing a dedicated visualization tools to support exploration of the SOS by combining existing multi-dimensional visualization techniques (Visual Data Mining, Visual Analytics) with narrative methods for collective sense-making to support stakeholders and policymakers to conceive, develop, and appraise alternative policies. Such tools will support the definition of storylines and pathways that take into consideration climate and socio-economic conditions, specifying the ones that generate performance failures, thus supporting the translation of complex models into actionable plans. Deliverable D5.3 "*Safe Operating Space visualization tool*" is expected to be available by end of September 2025.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

Results obtained from SOS-Water models are expected to follow common vocabulary related to water resources management and planning and therefore we do not foresee the need for additional mappings.

Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?

SOS-Water will include qualified references to other data when needed. For example in Zenodo, all publications and data that relate to each other are linked in their respective pages, and, in general, all SOS-Water material is collected under the same community page.

2.4 Increase data re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

The entire data management process is documented and verified; data collection methods, data editing, software used, analysis, codes, variables, standards, etc., so that it is possible to understand and reproduce the research. Every published set of data and codes will contain a readme-file to ensure ease of access. These readme-files will contain a list of relevant variable names, units, an outline of the folder structure, and a description of how codes and data work together.

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

The data used and generated is and will be freely and openly available through the selected institutional and public repositories, while complying with the obligations for data protection by the GDPR and any possible restrictions imposed by the data providers. We are licensing our data using the

standard reuse licenses as set out in the Grant Agreement. In general, the open data and metadata is made available for re-use under CC BY and CC0, respectively.

Will the data produced in the project be useable by third parties, in particular after the end of the project?

SOS-Water aims to guarantee open access as widely as possible during and after the project, while complying with the obligations for data protection by the GDPR and any possible restrictions imposed by the data providers.

Will the provenance of the data be thoroughly documented using the appropriate standards?

SOS-Water is following current best practices in data provenance standards to document the origin and lifecycle of data. The project is clearly documenting data sources, including who produced it, as well as data processing steps, and any changes made to the data over time, in order to assure transparency, reproducibility, and reusability. Information are provided through standardized metadata files in tabulated (CSV, Excel), raster (GRIB, NetCDF, GeoTIFF) or vector (Shapefiles, GeoJSON) format to meet the data provenance standards and facilitate data sharing. Examples of data and metadata already produced by the project can be found in the Zenodo SOS-Water community page.

Describe all relevant data quality assurance processes.

Data quality is ensured throughout the project by the establishment of a clear data management procedure and documentation for what concern data collection methods, data editing, software used, analysis, codes, variables, standards, etc. Standard protocols and formats are followed to guarantee high quality of data provenance and metadata documentations. Additionally, an internal review process led by the project coordinator and work package leaders has been set up at the initial stages of the project. This process rigorously evaluates all project outputs to ensure they meet the necessary quality standards before being published or submitted to scientific journals.

3. Other research outputs

SOS-Water aims to improve water systems and impact models to inform the computation of several environmental, social and economic indicators. As indicated above, those models are open-source (e.g., CWatM, PCR-GLOBWB 2.0) or are accessible promptly upon request (e.g., Jucar River Basin model). Codes of open-source models are available via hosting platforms such as GitHub to ensure continuous development and version control.

4. Allocation of resources

Costs for making data or other research outputs FAIR in SOS-Water

The possible costs for making data or other research outputs FAIR in SOS-Water are covered during the project by the project grant and after the project by partner institutions. For example, some of the scientific articles published so far charged OA publishing fees to the project. Additionally, some partners have institutional subscriptions that can support costs both during and after the project.

No cost is foreseen for depositing and accessing data in the identified institutional and public repositories, which all provide their services for free.

Responsible for data management in SOS-Water

Responsibility of data management in SOS-Water is of the coordinator of the project. The main contact person at each participant organization is responsible for the collection and transfer of data from its own research activities, within the framework of the DMP. After the end of the project, project partners are responsible for curating and preserving their data.

The DMP and Consortium Agreement (CA) outline the access rights of the consortium participants to data held by other consortium members and relevant ownership of data. The DMP and CA also identify responsibilities regarding the use of data during project lifetime as well as issues related to the long-term storage and maintenance and/or destruction of data after the project has been completed. Special attention will be paid to comply with the GDPR.

Long term preservation of data

We expect the data to remain available for the lifetime of the selected repositories. Costs and responsibility of long term preservation beyond project duration is the responsibility of the project partners.

5. Data security

Collected data is stored in the servers of respective partners and in the collaborative platform of the project on Microsoft SharePoint (if data size allows) and protected by defining authorised user groups and users and granting user IDs and passwords only to the members of the SOS-Water project. Partner organizations are taking care of the back-up and restore of the data by applying their internal back-up and restore policies. Data stored in the selected institutional and public repositories will be retained for the lifetime of the repositories.

6. Ethics

To guarantee the overall implementation of the respective ethical standards and the support in addressing new ethical issues that might arise during the project, the SOS-Water Steering Committee is responsible for monitoring all the ethical issues and report these annually to the General Assembly. Moreover, the SOS-Water project consortium appointed an independent Ethics advisor by month 3 of the project to ensure adherence to the Ethics issues concerning Human participation, Personal Data, and Artificial Intelligence. By month 15, an Ethics Advisory Board has been created with the purpose to facilitate communication between the Consortium and the Ethics advisor, and further support the monitoring of ethical issues throughout the project. The Ethics Board is composed by the Ethics Advisor, the project Coordinator and two other consortium members, with other members that might be added on a needing basis. A report prepared by the external independent Ethics Advisor is submitted as a deliverable at the end of each reporting period. If, ethical issues arise unexpectedly during the project, the SOS-Water consortium will immediately contact the Ethics Board and the Ethics Advisor and the granting authority to discuss and provide detailed information on the issue and how to handle it.

SOS-Water is carrying out research activities that require human participation such as workshops and stakeholder engagement processes. However, such activities only require the collection and handling of some personal data (mainly contact information such as email address/first name/last name/telephone number) and the participants' full and informed consent is ensured beforehand. The SOS-Water consortium ensures that personal data will be processed in accordance with ethical principles and GDPR that aim to limit the negative impact on the persons concerned and ensure fairness, transparency and accountability of the data processing, data quality and confidentiality. For the workshops organized in Vietnam, the consortium ensures compliance with the Vietnamese data

privacy law, “The Decree of Personal Data Protection” (PDPD), that also requires the collection of clear and valid consent before collecting personal data. Before receiving the consent of the participants, SOS-Water pays special attention to inform participants of all aspects of the research that are relevant to the participant's decision to participate, explain and clarify how information generated in the project are documented, stored, and shared, and that participation to interviews, surveys, workshops or meetings does not cause any harm to the participants. Participants will be provided with privacy notices that describe the aims, methods and implications of the research, storage and accessibility of the data, and how confidentiality is ensured. IIASA, the coordinator of SOS-Water, has a part-time (60%) data protection officer that can support and advise us in the compliance of personal data protection. The Informed Consent forms is further translated to participants’ language by the local partners. During surveys, interviews or workshops, the participant can choose not to participate and can withdraw at any stage without being penalized or disadvantaged.

The data collected adheres to the principles of data minimization, anonymization, and stringent data protection and security measures to ensure the safety of the participants and their privacy. Confidentiality and anonymity are ensured unless agreed differently with specific participants through signed agreements. Collected information is used for research purposes only and cannot be used for any other purpose without further consent (It will especially not be distributed to the media or for commercial purposes). Data are stored and processed in accordance with European best practice guidelines. Collected data is thus protected against unauthorized viewing, modifying, and erasing. Each partner organization in SOS-Water carrying out interviews, surveys or workshops is responsible for: (i) data security, (ii) handling the data in accordance with GDPR regulations, and (iii) their process and agreements related to data processing activities.

7. Other issues

The DMP ensures that data management and protection is compliant with EU principles and standards, and with relevant national data protection laws and institutional data management policies. Produced data will be treated according to the management guidelines and then implemented through relevant WPs.

The project also ensures ethical compliance when sharing data between EU and non-EU partners. The project complies with Vietnam data privacy law, “The Decree of Personal Data Protection” (PDPD) when collecting, sharing and storing personal data from Vietnamese stakeholders. Only basic personal data (e.g., name, affiliation, email address) from stakeholders are shared outside of Vietnam in the project’s collaborative platform on Microsoft SharePoint. Consent forms from participants are always collected beforehand (see also above). No data transfers from other project countries to Vietnam are foreseen.

For Switzerland and Norway, the EU Commission has considered their privacy and personal data legislation adequate, eliminating the need for additional guarantees when sharing or transferring data with these countries.

The SOS-Water consortium is committed to ethical research and to the principles described in the Charter of Fundamental Rights of the EU, the European Convention on Human Rights, and its supplementary Charter of Fundamental Rights of the European Union in Practice, as well as in pioneering documents (incl. Nuremberg Code, Declaration of Helsinki), and carries out its activities according to standard procedures as advised by the European Commission. More specifically, SOS-Water is subject to the European Code of Conduct for Research Integrity, as well as ethical standards in force by its academic and practice partners.

Deviation from Open-Access standards: For scientific peer-reviewed publications, partners prioritize the 'gold' model for open-access, i.e. publications are immediately available free of charge. In case

'gold' Open Access is unattainable, 'green' open access is pursued by submitting publications to institutional (e.g., IIASA-PURE, EAWAG-DORA, UPV-RiuNet) or public repositories (e.g., OpenAIRE Zenodo). All scientific papers published to date have been published with gold open access, but one, (Carlino et al., *Nature Sustainability*) which was published under a subscription-based model. This resulted from publication decisions taken during the submission process, without opting for Gold Open Access or immediate repository deposition at the time of acceptance. To address this deviation, the authors ensured compliance through 'green' open access by depositing the accepted manuscript version in an open repository (Zenodo: https://zenodo.org/records/17979596), in line with Horizon Europe requirements and publisher policies.

Disclaimer

Acknowledgement of funding

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101059264.