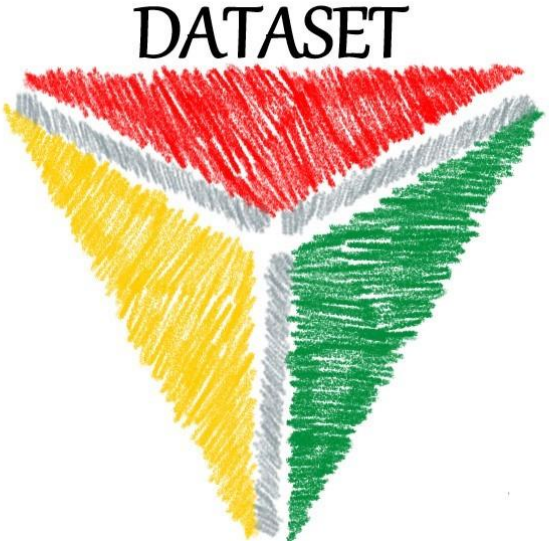

D0.2 – Data Management Plan

V 0.3

DATASET

“GrounDwater sAlinizaTion and pollution AsseSsmEnt Tool”

This project is part of the Water4All program supported by the European Union

Project number: WATER4ALL22_00084

Contractual due date: 07/2025

Author(s): Dawid Potrykus

Lead Beneficiary of Deliverable: GUT-VAN

Dissemination Level: Internal

Nature of the Deliverable: REPORT

Reviewer: Gianluigi Busico & Micòl Mastrocicco (internal members of DATASET, VAN)

History of changes

Version	Date	Author	Changes
1	**10/07/2025**	**D.P.**	
2	**15/07/2025**	**G.B.**	**Revision 1**
3	**28/07/2025**	**M.M.**	**Revision 2**
-	-	-	-

EXECUTIVE SUMMARY:

DMP is intended for use by DATASET project researchers only.

Distribution of the files in the database to external parties is forbidden.

Water4All Water security for the planet The DATASET project is supported by the European Union as part of Water4All

Table of Contents

1. PROJECT EXECUTIVE SUMMARY

DATASET (GrounDwater sAlinizaTion and pollution AsseSsmEnt Tool: a holistic approach for coastal areas) aims to develop an efficient tool to assist in the protection and regulation of water resources' use and management in coastal aquifers, considering all dynamic conditions caused by ongoing Climate Changes (CC) and more frequent Hydroclimatic Extreme Events (HEE).

The key objectives are:

1. Develop and apply a method to create vulnerability maps throughout a holistic assessment of coastal aquifers coupling the evaluation of agricultural leaching (AGL) and salinization processes (SP) in a single methodology to improve water governance aimed at early warning, prevention and mitigation of groundwater quality degradation in the present and future conditions, accounting for the impacts of CC and HEE.
2. Promote a paradigm shift in water management, through internet access to data and models, dissemination of management tools for planning and societal implication and their transfer to stakeholders.
3. Promote social awareness on the importance of groundwater and its sensitivity to CC and HEE from stakeholders and decision makers to end users.

The novelty of DATASET relies on the implementation of a holistic approach for the quantification of coastal areas' vulnerability to both SP and AGL. The DATASET index will be built on Open Data thanks to the possibility of using a growing number of available global and regional datasets, removing the dependency on local sampling and making the assessment completely cost-free. During the project, an open-source database named DATASET Open Database (DOD) was developed.

The DATASET index will be divided into two sub-indices: an Automatic DATASET Index (ADI) and an Improved DATASET index (IDI). ADI is a fully automated tool that enables effortless groundwater vulnerability assessment (GVA) for SP and AGL, based on open-source data and eliminating the need for manual input. IDI integrates both open-source and locally sourced data, providing operators with the flexibility to refine SP or AGL assessments when more detailed, site-specific data are available. The realization of a web-platform and of a completely new GIS-based plug-in will allow a wide audience to apply the DATASET index around the world, providing effective solutions not only for actual water resources management but even for future scenarios.

The DATASET methodology will be developed and calibrated in five pilot sites under consortium responsibility: Low Volturno River Plain (Italy), Po Delta Plain (Italy), Puck Bay (Poland), Vistula Spit (Poland), and Cape Flat (South Africa) through close collaboration with water utilities and stakeholders. This will enable immediate impacts on water security, aligned with EU directives and UN Sustainable Development Goals. The transnational consortium leverages complementary expertise to develop a transferable framework for developing GVA methods to AGL and SP, considering CC impact.

Partners:

- University of Campania “Luigi Vanvitelli” (VAN), Italy - Coordinating institution
- University of the Western Cape (UWC), South Africa
- Avignon Universitè (AU), France
- Gdańsk University of Technology (GUT), Poland
- Universidade de São Paulo (USP), Brazil

Work Package Leaders:

- VAN will lead WP0 on project coordination and WP3 on climate impacts assessment.
- GUT will lead WP1 on DATASET index creation.
- AU will lead WP2 on DATASET application development and calibration.

The partners will collaborate closely on integrating project activities with experimental data across the work packages.

2. DATA SUMMARY

As part of the DATASET project, various data and research outputs will be generated or processed. Their scope directly aligns with the project's objectives and the DATASET index methodology.

The developed DATASET Open Database (DOD) has been enriched with comprehensive information sourced from global and regional databases, following the findings of systematic reviews on SP and AGL (see D1.2: Report on AGL and SP). This will be raw data requiring processing to calculate selected vulnerability indices. Details and relevant notes (e.g., data provenance) will be outlined by the provision of a METADATA.txt file (see D1.1: DATASET Open database).

Data regarding the pilot and associated sites will be used for calibration and testing the new proposed vulnerability assessment method. It will consist of processed archival data, sourced from internal research, previous research projects, or provided by external entities. Processed using GIS software like ArcGIS Pro or QGIS. Such existing datasets are subject to conditions explicitly requested by the original data providers: that the received data not be commercialized or shared with third parties without prior written authorization; that the datasets be used exclusively for the stated research purposes; and that any publication or dissemination of the data cites its source.

Data types include:

- Global and regional dataset of hydro- and geological spatial information
- Thematic maps of pilot and associated sites
- Reports, guidelines, tutorials
- GIS plug-in
- Project WEBSITE

Knowledge (publications and data), software tools and maps generated by DATASET will be managed in accordance with the EU rules. Specifically, all manuscripts will be published under the Creative Commons Attribution (CC BY) license in “Gold Open Access” journal as main preference.

The data will be useful to researchers, water managers, policy makers, and the public sector.

The overall estimated data size is several Gigabytes across the consortium. A more precise estimation is not possible at this stage of the project. Possible updates will be included in subsequent versions of the DMP, which will be revised if needed.

All DMP relevant research assets DATASET is expected to generate can be tracked down to work packages and deliverables as listed in Table 1.

Tab.1. DMP relevant research assets of DATASET

WP No.	Deliverable / Milestone No.	Research asset description	Research asset ID*	Research asset format	Generated/ Reused
WP0	D0.1	Communication and Dissemination Plan	Do	.pdf	Generated
WP0	D0.2	Data Management Plan	Do	.pdf	Generated
WP0	D0.3	Web Platform	S	Web domain	Generated
WP0	M0.1	Final report on the dissemination and use of the DATSET index	R	.pdf	Generated
WP1	D1.1	DATASET Open Database (global and regional input files for GVA)	D	.geoTIFF, .shp	Reused
WP1	D1.2	Report on AGL and SP	R	.pdf	Generated
WP1	D1.3	Report on DATASET Index equation	R	.pdf	Generated
WP1	M1	Model and database implementation	T	.pdf	Generated
WP2	D2.1	Sites thematic raster maps of vulnerability	D	.geoTIFF	Generated
WP2	D2.2	Technical report on the definition of the model ratings and weights	R	.pdf	Generated
WP2	M2	Model user manual and tutorials	T	.pdf	Generated
WP3	D3.1	Report on future scenarios' prediction	R	.pdf	Generated
WP3	D3.1	Raster maps for sites according to future scenarios' prediction	D	.geoTIFF	Generated
WP3	D3.2	Management recommendations for stakeholders	Re	.pdf	Generated
WP3	D3.2	Raster maps for the pilot sites	D	.geoTIFF	Generated
WP3	M3	GIS plug-in	S	Python code	Generated
WP3	M3	Full operation of the DATASET index via the web platform	T	.pdf	Generated

*D: dataset/database, S: software/app, Do: documents/datasets for internal use, R: report, T: tutorial/guidelines, Re: recommendations

3. FAIR DATA

Following the principles of FAIR data, all DATASET project research outputs will be findable, accessible, interoperable, and reusable. Knowledge (publications and data), software tools and maps generated by DATASET will be managed in accordance with the EU rules. They will be shared as early and openly as possible providing guidance for potentially interested users.

3.1. MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA

All DATASET DMP-relevant research assets, including reports, deliverable, manuscript, links and codes, will be made available in the project website at the link https://www.datasetw4all.info/. The DATASET Open Database (DOD) will be stored in the INTERNXT repository https://internxt.com, which is linked to the project website. The prepared publications associated with the DATASET project's outcomes will feature links to the corresponding data repositories.

For all global and regional data placed in the DATASET Open Database (DOD), a METADATA.txt file has been prepared that provides essential information about how data in a database are structured, organized, and characterized. It acts as a blueprint for efficient data retrieval, integration, and analysis by detailing data types, relationships, constraints, and indexing. Properly managed metadata ensures data consistency, improves query performance, and enhances data governance by providing greater control over data access and usage. This foundational layer supports the overall integrity and usability of the database. Essentially, metadata answers questions like who created the data, when it was created, what format it is in, and how it should be used, making it an essential component for data management and interpretation. Accordingly, dedicated metadata has been created for each stored file within the common database, following the structure below:

- FILE NAME: [Insert the name of the stored file]
- FILE FORMAT: [Specify if vectorial, raster, or table]
- FILE RESOLUTION AND EXTENT: [Clarify the spatial extent of the file, e.g., regional, national, or global]
- FILE COORDINATE SYSTEM: [Indicate the Coordinate system with the associated DATUM, e.g., WGS 84, EPSG:4326]
- ATTRIBUTE LABEL DEFINITION: [Provide a description of pixel values or attribute table content]
- SOURCE LINK FOR DOWNLOAD: [Insert the direct link to the original resources if available]
- GENERAL INFO: [Provide a brief description of the data sources, how it was created, how it can be used, and any additional information that can help in better understanding how to correctly use the data]
- REFERENCES: [List of source citations for the database, such as manuscripts, repositories, etc.]

Example of Metadata information:

FILE NAME: GLOBALWTDFTP

FILE FORMAT: NetCDF extracted in GeoTIFF (raster)

FILE RESOLUTION AND EXTENT: 30" global grid and all files are organized by continents.

FILE COORDINATE SYSTEM: GCS WGS84

ATTRIBUTE LABEL DEFINITION: X Latitude (°), Y Longitude (°), Z Groundwater table (m)

SOURCE LINK FOR DOWNLOAD:

http://thredds-gfnl.usc.es/thredds/catalog/GLOBALWTDFTP/catalog.html

GENERAL INFO:

The data offer a simulated estimation of the stationary water table depth for the entire world at ~250 m planar resolution. Data are organized in monthly means/ (over 10yrs of model run) and annual means. All files are organized by continents in NetCDF which can be read by ArcGIS version 9 or later. The model coupled the vertical soil water balance with groundwater recharge and discharge, water table rise and fall, and adaptive plant root uptake depth to meet ET demand inferred from satellite observations of Leaf Area Index. It is worth stressing that the high-resolution depth to water map has been calibrated and validated using measures of water table coming from more than 30 k monitoring stations located all over the continent (Fan et al., 2013).

REFERENCES

Fan, Y., H. Li, G. Miguez-Macho (2013) Global patterns of groundwater table depth, Science, 339 (6122): 940-943, doi:10.1126/science.1229881

Fan, Y., G. Miguez-Macho, E.G. Jobbágy, R.B. Jackson, C. Otero-Casal (2017) Hydrologic regulation of plant rooting depth, Proceedings of the National Academy of Sciences of the United States of America, Vol 114, No 40, 10572–10577, doi: 10.1073/pnas.1712381114.

3.2. MAKING DATA ACCESSIBLE

An internal exchange platform was created to facilitate data storage and sharing among project partners. As a temporary repository for files, official communications, and video calls, the Microsoft Teams platform is being used and hosted by the University of Campania “Luigi Vanvitelli”. A dedicated channel named "DATASET project" has been created, featuring an internal storage system, which will be available for the entire duration of the project. Access to the "DATASET project" channel is closed and intended for use only by researchers involved in the DATASET project.

The final, permanent storage location for the DATASET project's processed data (DATASET Open Database) for open dissemination will be the INTERNXT service linked to the project website. This solution provides data storage with lifetime access (long-term preservation). Some of the data has been uploaded to the database, while for larger files, download links were provided. Access to the DATASET Open Database (DOD) via the INTERNXT service is available after users’ registration on the project's website.

Other research outputs (e.g., reports, tutorials, guidelines) will be made available on the project website.

3.3. MAKING DATA INTEROPERABLE

Whenever possible, data will follow open formats and standards from relevant hydrological communities. Spatial data stored in DATASET Open Database (DOD) can be easily opened using GIS software like ArcGIS or QGIS. Specifically, DATASET Open Database (DOD) will contain files in various formats, such as raster (e.g., GeoTIFF), vectorial (e.g., shapefiles) or table (e.g., XLS).

In all other cases (e.g., reports, documents) data will be stored in PDF-format files supported by any open-source software. Terminologies will be aligned with commonly used ontologies; where necessary, project-specific vocabularies will be documented and mapped.

3.4. INCREASE DATA RE-USE

All main project’s outcomes will be licensed under Creative Commons (e.g., CC BY 4.0). Provenance will be tracked, and data quality will be ensured through validation procedures and cross-checks with field observations.

4. OTHER RESEARCH OUTPUTS

Other outputs include software (e.g., GIS plug-in), workflows, protocols, and decision-support tools. These will be managed according to the FAIR principles. Software will be open-source where possible and hosted on project website or open repository (e.g., Zenodo) with appropriate documentation and licensing.

5. ALLOCATION OF RESOURCES

Data management activities (storage, curation, sharing) are included in the project budget. Each WP leader and project partner is responsible for specific datasets.

For the DATASET project, an initial exchange platform was created to facilitate data storage and sharing among project partners. As a temporary repository for files, official communications, and video calls, the Microsoft Teams platform is being used. A dedicated channel named "DATASET project" has been created, featuring an internal storage system, which will be available for the entire duration of the project. The final,

permanent storage location for the DATASET project's processed data, for open data dissemination, will be the INTERNXT service, which provides data storage with lifetime access (long-term preservation).

6. DATA SECURITY

All data will be stored securely with routine backups. Sensitive data will be encrypted and access controlled. Disaster recovery protocols will be in place at each partner institution.

Access to the DATASET Open Database (DOD) via the INTERNXT service is available after users' registration on the project's website. The INTERNXT service ensures appropriate data security and offers comprehensive disaster recovery protocols, including automated backups, ransomware protection, and secure data storage thanks to the AES-256 zero-knowledge encryption. Client-side end-to-end encryption and post-quantum cryptography are implemented to ensure data security even in the face of future threats. Moreover, the source code is made public on GitHub and can be reviewed, inspected, and verified personally by anyone.

7. ETHICS

No ethical issues are foreseen within the DATASET project. Some sensitive environmental data may be restricted by the original providers.

The project plans to utilize Artificial Intelligence (AI). However, the use of AI algorithms will not raise ethical concerns regarding human rights and values. This is because the application focuses on physical and chemical data related to groundwater resources. For example, the AI algorithms will be deployed to analyse data such as nitrate concentration or rainfall.

The ethical aspects of the Partnership's activities are governed by the Consortium Agreement, which incorporates the binding regulations of the EU Charter of Fundamental Rights and the EU Code of Ethical Research.

8. OTHER ISSUES

The project will adhere to the University of Campania "Luigi Vanvitelli" and other partner institutions data management procedures, including compliance with GDPR where applicable.