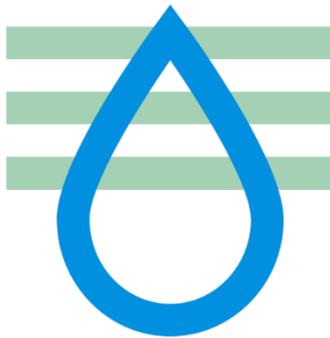


DELIVERABLE D1.1
DATA MANAGEMENT PLAN



eGROUND WATER

Citizen science and ICT-based enhanced information systems for groundwater assessment, modelling and sustainable participatory management (GA n. 1921).

DELIVERABLE D1.1	DATA MANAGEMENT PLAN
Related WP	WP1
Leader	UPV
Authors	Manuel Pulido-Velazquez, Hector Macian-Sorribes, Adria Rubio-Martin, Esther Lopez-Perez, Luis Nunes, Juan Caravaca
Contact	mapuve@upv.es
GA number	1921
Instrument	PRIMA
Project start date	2020-03-01
Project duration	48 months
Website	
Abstract	<p>This deliverable describes the data management life cycle of eGROUNDWATER, including a summary of the data collected, pre-processed and produced by the project, the provisions made to ensure FAIR (findable, accessible, interoperable and reproducible), the resources that eGROUNDWATER will allocate to ensure FAIR data, the data security guidelines followed by the project, the ethical aspects of eGROUNDWATER data and a table describing all the datasets that will be generated. This deliverable complies with the DMP template set by H2020, and it will be regularly updated during the project lifetime.</p>

DISSEMINATION LEVEL

<input checked="" type="checkbox"/>	PU	Public, fully open
<input type="checkbox"/>	CO	Confidential, restricted under conditions set out in Model Grant Agreement
<input type="checkbox"/>	CI	Classified, information as referred to in Commission Decision 2001/844/EC

VERSIONING AND CONTRIBUTING HISTORY

VERSION	DATE	MODIFIED BY	MODIFICATION REASONS
V0.01	2020-07-02	Manuel Pulido-Velazquez, Hector Macian-Sorribes, Adria Rubio-Martin, Esther Lopez-Perez	Generation of the first draft
V0.02	2020-07-20	Luis Nunes	Input from UAlg
V0.03	2020-07-21	Juan Caravaca	Input from VNA

TABLE OF CONTENTS

1	Data summary	5
1.1	Data collection and generation.....	5
1.2	Types and formats of data	6
1.3	Re-use of existing data.....	8
1.4	Origin of data	8
1.5	Size of data.....	9
2	FAIR data	10
2.1	Making data findable	10
2.1.1	Data Management Process.....	10
2.1.2	Naming conventions and keywords.....	11
2.1.3	Version numbers.....	12
2.2	Making data openly accessible	12
2.3	Making data interoperable	13
2.4	Increase data re-use (through clarifying licenses).....	13
3	Allocation of resources.....	14
4	Data security.....	15
4.1	Data made openly accessible through Zenodo	15
4.2	Data stored in the EIS.....	16
5	Ethical aspects	16
5.1	Informed Consent and Assent.....	17
5.2	Data privacy	18
6	eGROUNDWATER Data Record of collected and generated datasets	19

LIST OF FIGURES

Fig 1.	Information flows in eGROUNDWATER's EIS.....	5
Fig 2.	Data management process of eGROUNDWATER	10



1 Data summary

1.1 Data collection and generation

The collection, post-processing and exchange of data is at the very heart of eGROUNDWATER, since the project will combine data collected using citizen science with data from traditional (e.g. piezometers) and innovative (e.g. remote sensing) information systems to develop Enhanced Information Systems (EIS, Fig 1). The purpose of data collection of eGROUNDWATER is to load it into the EIS in order to post-process and combine them to provide enhanced knowledge of groundwater bodies and thus advance towards a more efficient, transparent and resilient groundwater management. It will also be used as input for the groundwater models and decision-support systems (DSS) to be developed by eGROUNDWATER. After the development of those models and DSS, the EIS will collect the management decisions on the groundwater body and will forward it to the groundwater users.

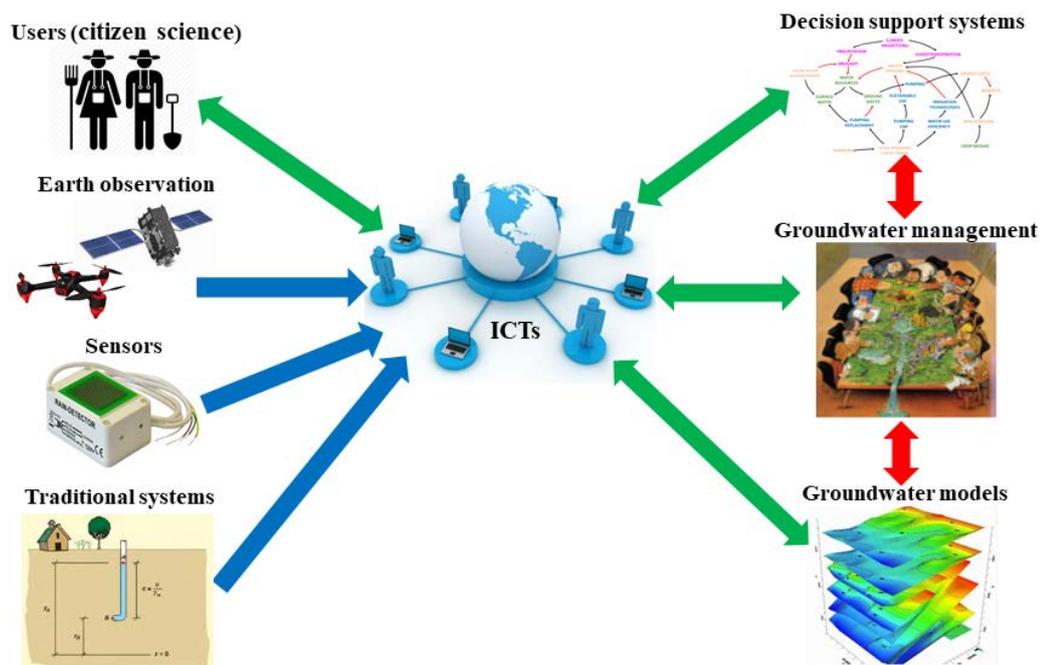


Fig 1. Information flows in eGROUNDWATER's EIS

Data collection is related with the overall objective of the project (support sustainable participatory groundwater management in the Mediterranean regions through the design, testing and assessment of enhanced information systems) and with specific objectives 2 (To develop innovative ICT-based tools to enhance information acquisition using citizen science techniques, tailored to four case studies), 3 (To assess the impact of enhanced information availability in groundwater assessment and modelling in four case studies: the project will determine how increased knowledge availability will reduce uncertainty on the estimation of groundwater cycle components by improving existing modelling tools and enabling the use of new models) and 4 (To develop improved participatory groundwater management strategies, supported by the enhanced information system

DELIVERABLE D1.1: Data Management Plan

(EIS) and groundwater flow modelling, and management tools for sustainable and efficient groundwater use).

Data generated by eGROUNDWATER will include:

- Results of the post-processing of collected data (e.g. soil moisture maps from remote sensing, groundwater head maps combining piezometers and user-made measurements on well depth).
- Results of the models and decision-support systems improved or developed by eGROUNDWATER, fed by the data collected and the post-processed data.

eGROUNDWATER's generated data will expand and homogenize the information collected from all the data sources and obtained from the models and decision support systems to provide a clear and comprehensive picture of the groundwater body and thus facilitate decision-making. Data generation is related with the overall objective of the project (support sustainable participatory groundwater management in the Mediterranean regions through the design, testing and assessment of enhanced information systems) and with specific objectives 2 (To develop innovative ICT-based tools to enhance information acquisition using citizen science techniques, tailored to four case studies), 3 (To assess the impact of enhanced information availability in groundwater assessment and modelling in four case studies: the project will determine how increased knowledge availability will reduce uncertainty on the estimation of groundwater cycle components by improving existing modelling tools and enabling the use of new models) and 4 (To develop improved participatory groundwater management strategies, supported by the enhanced information system (EIS) and groundwater flow modelling, and management tools for sustainable and efficient groundwater use).

1.2 Types and formats of data

eGROUNDWATER will collect and generate data in three different types:

- Point time series: they will be attached to a particular location (e.g. pumping well, piezometer, spring) and cover a particular time span (between the beginning and the end of the recording period) with a certain periodicity (e.g. daily, monthly, yearly).
- Gridded time series: they will refer to an area (e.g. aquifer, irrigation district) that will be divided into grid points through a mesh, and cover a particular time span (between the beginning and the end of the recording period) with a certain periodicity (e.g. daily, monthly, yearly).
- Images: they will refer to an area (e.g. aquifer, irrigation district) that will be covered with a single or with a collection of images, and to a particular time (year, month, day and hour). In eGROUNDWATER, image data will come mainly from remote sensing using drones or satellites.

The data formats employed by eGROUNDWATER will depend on the type of data.

1. Point time series

- a. *Text files* (.txt, .csv) separated by spaces or semicolons will be the standard format for these data. These formats are easily created and read by any programming language (e.g. Python, C++, Java, R, Visual Basic) or software (e.g. Notepad++, Microsoft Excel, OpenOffice).



- b. *Microsoft Excel files* (.xls, .xlsx, .xlsm) will be used in case that the text file format was not possible. It offers similar capabilities than text files, but the resulting files are usually larger in terms of size, and it is bound to particular software products (Microsoft Excel, OpenOffice).

2. Gridded time series:

- a. *NETCDF* (.nc) will be the standard format for these data, in particular in its NETCDF4 version. This format efficiently packs all the gridded time series of a particular variable or collection of variables into a single archive, which can be easily generated and read using GIS servers (e.g. QGIS, ArcGIS) or programming languages (e.g. Python, R). It is a standard format for the provision of meteorological data and climate change scenarios.
- b. *GRIB* (.grib) will be used in case that NETCDF file format was not possible. This format is a standard format for the provision of meteorological data used by the WMO among others. GRIB files are usually more efficient than NETCDF in terms of size, but creating and reading a GRIB file is distinctly heavier than NETCDF.
- c. *GEOTIFF* (.tiff) will be used in case that the previous file format were not possible. In this case, a collection of files will be generated, each one attached to a particular time (year, month and day). It offers the easiest alternative in terms of creation and reading, but the packing of the information is distinctly cumbersome than the previous ones and will demand more space.

3. Images from drones and satellites

- a. *JPG* (.jpg) is a standard format in cameras installed in drones.
- b. *DNG* (.dng) is also standard in cameras installed in drones.
- c. *GEOTIFF* (.tiff) will be the standard format for remote sensing images after being processed. This standard format is composed by a tiff image which is geo-located, thus allowing to know the exact position of each pixel. The CRS will be 4326 preferably.
- d. *JP2* (.jp2) will be the standard format for the original remote sensing images collected by Sentinel-2 program by ESA.

4. GIS data

- a. *ESRI shapefile* (.shp) will be the primary format for vectorial information (points, lines, polygons). Each shapefile will be associated to a series of additional information files such as .dbf (database) and .prj (cartographic projection).
- b. *Formats included under “gridded time series”* are available for use in GIS.
- c. *GIS raster formats* (.dat, .grd) will be employed as an alternative to gridded time series formats.
- d. *Raster image files* (e.g. .bmp, .tif, .gif, .jpeg) will be employed in rasterized photos.
- e. *Supplementary files* (e.g. metadata) will be provided in additional formats as .xml.

1.3 Re-use of existing data

eGROUNDWATER relies on some data sources (piezometers, satellite images) that are already in place and have produced data from years or decades ago. The project will also use outputs from simulation models obtained in previous projects, which will help estimating values where direct measurement is not possible, and to forecast. This data from the past will be used and combined for the calibration of the groundwater models and decision support systems that will be integrated into the EIS. The treatment of existing data from these sources will follow the same process as done for the new data, outlines previously.

1.4 Origin of data

The EIS foreseen by eGROUNDWATER will combine information from different origins. As indicated in Fig 1, the origins of the data sets are grouped on four different categories according to the source of the information: users, earth observation, sensors and traditional systems. For each one, the origin of the data will be:

1. Users:

- a. *Groundwater levels*: measured by groundwater users as depth measurements at boreholes.
- b. *Pumping rates*: measured by groundwater users per borehole or agricultural plot.
- c. *Water quality*: measured by groundwater users or third people (depending on the case study and the stage of the project) using portable kits.

2. Earth observation technologies:

- a. *Meteorological variables*: obtained from the ERA5 reanalysis dataset through the Climate Data Store of the Copernicus Climate Change Service (C3S) <https://cds.climate.copernicus.eu/cdsapp#!/home>.
- b. *Crop evapotranspiration*: obtained from satellite images from the satellite programs Sentinel, Pleiades and Landsat, through the Earth Observing System (<https://eos.com/>) or ONDA DIAS (<https://www.onda-dias.eu/cms/data/catalogue/>), or using drone flights (directly obtained from the drone pilot and the end of each flight).
- c. *Vegetation dynamics*: using the same sources of information as the crop evapotranspiration.

3. Sensors:

- a. *Soil moisture*: obtained using frequency domain reflectometry (FDR) sensors.
- b. *Evapotranspiration*: obtained using eddy covariance and surface renewal sensors.
- c. *Groundwater levels*: measured using data loggers (integrated sensor and datalogger for measurement of water level, temperature, and conductivity in groundwater, CTD).
- d. *Pumping rates*: measured using data loggers (CTD and portable water level meter).



4. Traditional systems:

- a. *Meteorological variables*: obtained from weather stations from the meteorological agencies, river basin authorities and farmer associations. In case they are available, meteorological variables will be obtained through regional, national and international gridded time series (such as the Spain02 project for the Spanish territory, <http://www.meteo.unican.es/es/datasets/spain02>).
- b. *Groundwater levels*: obtained from piezometers from the river basin authorities, national environmental online databases, geological surveys and farmer associations.
- c. *Groundwater quality*: measured from groundwater quality measurement wells from the river basin authorities, national environmental online databases, geological surveys and farmer associations.
- d. *Groundwater discharge*: it will be measured in an indirect way, through measuring the discharge from springs or from streamflows in river reaches that are known to share a connection with the aquifer. These measurements will come from river basin authorities, national environmental online databases, regional and national governments and user associations.

1.5 Size of data

The exact size of the data cannot be anticipated at the current stage, since it will depend on the particular configuration of each EIS and the number of sources and origins combined (e.g. number of users introducing data, number of sensors deployed, number of satellite images used). The expected size of the data is in the order of Tb, given the large amounts of data that are expected to be collected.

To provide an overview of the size of the data, the following examples of the size of individual dataset are provided:

- 1 .txt file with 180 records on a particular point (15 years of monthly data): 3 KB.
- 1 .xls file with 180 records on a particular point (15 years of monthly data): 24 KB.
- 1 .xlsx file with 180 records on a particular point (15 years of monthly data): 19 KB.
- 1 .nc file with 65 years of daily precipitation and temperature records covering Spain with a regular mesh of ~10 km: 1.3 GB.
- 1 drone image: 2 MB (.jpg) or 20 Mb (.dng).
- 1 GIS project file (.prj): 5 MB
- 1 raster image (.bmp, .tig, .gif, .jpeg): up to 50 MB depending on the case study, the area covered and the resolution of the image.
- 1 .tiff (geotiff) of an area of 10.000 square kilometers (1 picture from Sentinel 2 satellite): 102MB

The size of the data collected will be regularly annotated and reported in the updated versions of the data management plan.

2 FAIR data

2.1 Making data findable

2.1.1 Data Management Process

In compliance with article 29.3 of the Grant Agreement, the management of the data collected and produced by eGROUNDWATER will follow the process presented in Fig 2.

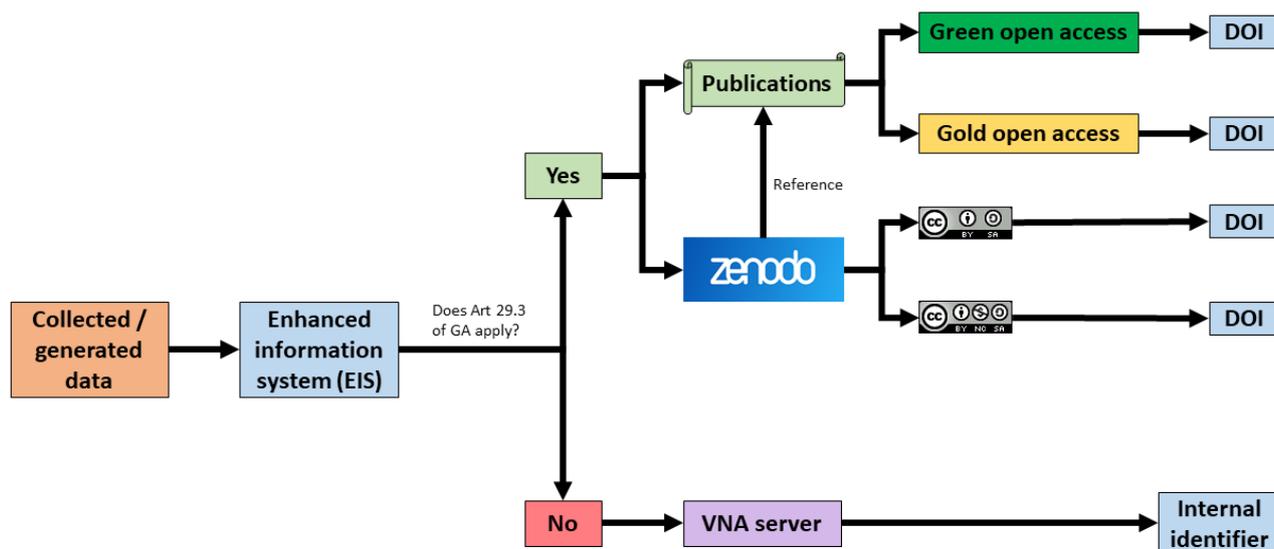


Fig 2. Data management process of eGROUNDWATER

All the data collected and generated by eGROUNDWATER will be loaded into the EIS. For each data set or subset, the key management decision will be the application or not of Article 29.3 of the Grant Agreement (GA). This article indicates that partners should, as soon as possible, deposit in a research data repository all the data, including associated metadata, needed to validate the results presented in scientific publications, and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user.

Consequently, the data that complies the requirements of Art 29.3 of the GA will be deposited in the Zenodo public repository (an OpenAIRE and CERN collaboration, <https://www.zenodo.org/>). Zenodo assigns a Digital Object Identifier (DOI) and a Creative Commons (CC) 4.0 license (<https://creativecommons.org/licenses/>). This DOI will be properly referenced in the publication(s) that use the deposited data. Data deposited in Zenodo during eGROUNDWATER will be shared under the Creative Commons ShareAlike (CC-BY-SA) and Creative Commons NonCommercial ShareAlike (CC-BY-NC-SA). The ShareAlike licenses bound the user to share any product using the data following the same license as the data used, while NonCommercial licenses prevent any user to make commercial use of the shared data.

If Art 29.3 was not applicable, data will be stored in a server of VisualNacert, which has a long-term experience on data collection and processing, complying will all legal requirements concerning data



protection measures. An internal identifier will be assigned to this data using VisualNAcert standards. Access to these datasets by the rest of the partners will be granted under the conditions set by eGROUNDWATER's Consortium Agreement.

Regardless of the Data Management Process, any dataset collected or generated by eGROUNDWATER will be properly annotated in the eGROUNDWATER Data Record (section 6 of this document), including the partner responsible of the collection/generation and the partner responsible of storing it.

2.1.2 Naming conventions and keywords

eGROUNDWATER will apply the following naming conventions to all the datasets generated, linked using underscores ("_").

1. Prefix "eGW", indicating that is data produced by eGROUNDWATER
2. Data origin:
 - a. "Coll" if the data is collected.
 - b. "Gen" if the data is generated.
3. Name of the case study:
 - a. "Tim" for the oasis archipelago in the Algerian Sahara near Timimoun, region of Adrar (Algeria).
 - b. "AiT" for the Ain Timguenay aquifer (Morocco).
 - c. "CdF" for the Campina de Faro aquifer (Portugal).
 - d. "RqU" for the Requena-Utiel aquifer (Spain).
4. Origin of the data:
 - a. "Cit" for citizen science.
 - b. "EOT" for Earth observation technologies.
 - c. "Sen" for sensors.
 - d. "Trad" for traditional systems.
5. Name of variable:
 - a. "GwL" for groundwater levels.
 - b. "PRt" for pumping rates.
 - c. "WQ" for water quality (in this case followed by the abbreviation of the pollutant)
 - d. "Met" for meteorological variables (in this case followed by the abbreviation of the variable).
 - e. "SMo" for soil moisture.
 - f. "GwD" for groundwater discharge (followed by "SP" if it is measured in a spring or by "Int" if it is measured through stream-aquifer interaction).
 - g. "CrET" for crop evapotranspiration (followed by 0 if it is the reference evapotranspiration).
 - h. "VD" vegetation dynamics (in this case followed by the abbreviation of the variable)
6. Type of variable:
 - a. "PTS" for point time series.
 - b. "GTS" for gridded time series.
 - c. "Img" for images.
 - d. "GIS" for GIS data

DELIVERABLE D1.1: Data Management Plan

7. Suffix “VX_Y” indicating the version of the data (e.g. the version 1.0 would be labelled as V1_0).

For example, groundwater level data collected by on user in the Ain Timguenay area aquifer would be labelled as *eGW_Coll_AiT_CiT_GwL_PTS_V1_0*.

To optimize the possibilities of data re-use by further users interested in any of the eGROUNDWATER case studies and variables, the datasets deposited in Zenodo will be given at least four keywords that will refer, respectively, to the project (eGROUNDWATER), the case study (e.g. Requena-Utiel), the variable (e.g. pumping rates) and the origin (e.g. citizen science). The partner that deposits the data can add more keywords at its own convenience.

2.1.3 Version numbers

eGROUNDWATER datasets will be given version numbers with the following convention:

- Data collected and generated during setting up the EIS and their initial components (e.g. data collected during the setup of the collection processes or generated during the calibration and validation of the models included in the EIS) will be given the number 0, and each update until they are validated will be given consecutive numbers using two decimal positions starting in version 0.01 (e.g. version 0.02 would indicate the second trial version, while 0.10 would indicate the tenth trial version).
- The first version after the setup of the EIS will be given the number 1.00. From this point, any minor update in any collection method or any model included in the EIS will increase the number in the decimal positions for all the datasets used by the EIS (for example, the first update will be given the number 1.01, and any minor change in a component of the EIS will increase all the datasets by 0.01 even though they were not changed). This homogenization of versions will ease the association of different data sets to the version of the EIS to which they relate.
- Any major update in the EIS, the collection methods or the models (including their entire replacement) will increase the number by 1 (e.g. the version after the first major update will be given the version number 2.00).
- In case of data related to collected variables or models added after the first version of the EIS (e.g. version 1.00 of the EIS does not include water quality, but it is added in version 2.00), data collected before the validation of the new data or model will be given a version starting by 0, but after validation they will access directly the current version of the EIS (e.g. the data associated with a new variable that will be included in version 3.00 of the EIS will be given numbers in the form 0.XX before the validation of the collection process of this variable, but once it is validated and included in the EIS they will directly jump to version 3.00, according to the version of the EIS to which they refer).

2.2 Making data openly accessible

As indicated by Fig 2, in compliance with Art 29.3 of the GA, the collected and generated data made openly available to the general public by default will be the data required to validate the results presented in eGROUNDWATER’s scientific publications. Since they may include the use of personal data, or data that could be used to obtain information on personal or professional behavior of



people, these datasets will be shared under restrictions aiming at protecting the privacy and professional interests of the users involved in eGROUNDWATER (see section 5). This data will be accessed through the Zenodo repository, being accessible by the internet with no additional tools than the ones required to visualize the datasets.

Furthermore, users contributing to the EIS by providing their own measurements through citizen science will be given full access for free to the datasets they provided during and after the end of the project, as long as they keep their status of contributing users. For this purpose, a list of contributing users to the EIS will be created, maintained and updated during and after the end of the project. These datasets will be accessible through the EIS, into which the users will access, using their personal account (with username and password), by the internet or a cellphone application. The EIS will implement adequate visualization techniques depending on the type of data provided by the users, who will be properly instructed on the use of the EIS.

Apart from the previous situations, open access to the data collected and generated by eGROUNDWATER will be restricted to the conditions of not hindering eGROUNDWATER's obligations on protection (Art. 27 GA) and exploitation (Art 28 GA) of results. The decision on granting open access to the datasets collected and generated will be made by the Steering Committee, after verifying that the access granted does not collide with the Exploitation Plan (D7.4). The Steering Committee will grant open access to all the datasets generated or to a particular subset of data using two mechanisms:

- Open access to the general public: which will be materialized through uploading the datasets into the Zenodo repository.
- Open access to particular users: which will be materialized through an EIS account given to the users to who the access is granted. The datasets to which access will be granted will be linked to the type of account with which the users will access the EIS.

2.3 Making data interoperable

All the datasets collected and generated by eGROUNDWATER will be included in the Data Record (section 6), which will give information on the formats, units and conventions used. Unless specified otherwise, eGROUNDWATER will follow the International System of Units and the standard conventions related to the particular data formats if any (e.g. NETCDF .nc files will follow the NETCDF4 CF 1.8 convention, <http://cfconventions.org/>).

In case of using own conventions, the datasets affected by them will be indicated in the Data Record and the specifications of those conventions will be included in this document to guarantee the interoperability of the related datasets.

2.4 Increase data re-use (through clarifying licenses)

In order to maximize the reuse of the openly accessible data of eGROUNDWATER, Creative Commons (CC) 4.0 licenses (<https://creativecommons.org/licenses/>) will be used in all the datasets

DELIVERABLE D1.1: Data Management Plan

shared unless agreed otherwise by the Steering Committee. In particular, the following CC licenses will be used:

- Creative Commons ShareAlike (CC-BY-SA): This license lets others remix, adapt, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to “copyleft” free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.
- Creative Commons NonCommercial ShareAlike (CC-BY-NC-SA): This license lets others remix, adapt, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.

In case that the data is shared using Zenodo, the repository allows the user to introduce the desired Creative Commons License and it is automatically shown to the users accessing these datasets. In case the data is shared using the EIS, the user will be properly informed on the licenses used in the first access to the platform.

Data made openly accessible to the users contributing to the EIS will be made accessible to them immediately after its introduction into the platform, and will remain accessible and re-usable as long as they hold the status of contributors. Data made openly accessible through Zenodo will remain accessible and re-usable under the following conditions:

- The data made openly accessible in compliance of Art 29.3 of the GA will be made accessible and available for re-use the day in which the first version of the related manuscript is sent to the scientific journal, and will be permanently available in Zenodo. In the unlikely event that Zenodo stops its operation, the corresponding author of the paper will upload the data into another repository with features comparable to Zenodo.
- The data made openly accessible by decision of the Steering Committee will be made accessible and available for re-use for the period established by it. At the end of the project, the Steering Committee will appoint, from the project PIs and the Exploitation Manager, and for each data set or subset, one or several data curators that will have the same responsibilities as the Steering Committee for the project afterlife.

3 Allocation of resources

The costs for making data FAIR in eGROUNDWATER will consist mainly on the personnel costs associated with post-processing the data collected and generated to adapt it to the formats and conventions set in this Data Management Plan. Furthermore, there will be personnel costs associated with giving open access to the EIS contributors so they can access the data they provided, granting open access to particular users through the EIS, as well as the personnel costs associated with uploading data into the Zenodo repository in case of data complying with Art 29.3 of the GA or data shared to the general public as mandated by the Steering Committee.



The Direct Personnel Costs item of eGROUNDWATER's budget already takes into account these costs. In any case, the staff effort required to make data FAIR is not considered significant compared to the global Direct Personnel Costs. Given that the Zenodo repository has no costs associated with depositing data, there will be no other costs associated with making data FAIR.

During the project, the Steering Committee (formed by the PIs and the Exploitation Manager) will be the main responsible for data management, which will follow the guidelines of this Data Management Plan. In any case, decisions on data management will not be in conflict with the exploitation of the project's results foreseen by the Exploitation Plan (D7.4). Its members will be responsible of informing their respective research teams on the provisions set by this Data Management Plan and monitoring its compliance in their own institution.

After the project end, if not specified by the Exploitation Plan (D7.4), the Steering Committee will appoint, from the project PIs and the Exploitation Manager, and for each data set or subset, one or several data curators that will have the same responsibilities as the Steering Committee for the project afterlife.

4 Data security

4.1 Data made openly accessible through Zenodo

The data shared openly using the Zenodo repository will be subject to the provisions made by Zenodo (<https://about.zenodo.org/policies/> and <https://about.zenodo.org/infrastructure/>):

- Zenodo security guidelines:
 - CERN Data Centre: the Data Centre used by Zenodo is located on CERN premises and all physical access is restricted to a limited number of staff with appropriate training and who have been granted access in line with their professional duties. Zenodo staff do not have physical access to the CERN Data Centre).
 - Servers: Zenodo servers are managed according to the CERN Security Baseline for Servers, meaning e.g. remote access to them are restricted to Zenodo staff with appropriate training, and the operating system and installed applications are kept updated with latest security patches via our automatic configuration management system Puppet.
 - Network: CERN Security Team runs both host and network based intrusion detection systems and monitors the traffic flow, pattern and contents into and out of CERN networks in order to detect attacks. All access to zenodo.org happens over HTTPS, except for static documentation pages which are hosted on GitHub Pages.
 - Data: Zenodo stores user passwords using strong cryptographic password hashing algorithms (currently PBKDF2+SHA512). Users' access tokens to GitHub and ORCID are stored encrypted and can only be decrypted with the application's secret key.
 - Application: they are employing a suite of techniques to protect the sessions from being stolen by an attacker when logged in and run vulnerability scans against the application.
 - Staff: CERN staff with access to user data operate under CERN Operational Circular no. 5, meaning among other things that:

DELIVERABLE D1.1: Data Management Plan

- staff should not exchange among themselves information acquired unless it is expressly required for the execution of their duties.
- access to user data must always be consistent with the professional duties and only permitted for resolution of problems, detection of security issues, monitoring of resources and similar.
- staff are liable for damage resulting from any infringement and can have access withdrawn and/or be subject to disciplinary or legal proceedings depending on seriousness of the infringement.
- Replicas: All data files are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
- Zenodo retention period: Items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.

4.2 Data stored in the EIS

The data not shared openly using Zenodo will be stored in the servers of VisualNAcert, which will host the EIS and its functioning.

- VisualNAcert data security guidelines: personal data will be dissociated from non-personal data by storing it on different spaces. Regarding data access security, VisualNAcert has several cloud providers such as Amazon Web Services or OVH. More information regarding data security could be obtained from our providers (<https://aws.amazon.com/security/> and <https://www.ovh.co.uk/personal-data-protection/security.xml>)

5 Ethical aspects

The main ethical issue of eGROUNDWATER concerning data refers to the use and processing of personal data; since the EIS will collect groundwater variables associated to specific users. Some of these datasets may be considered as tracking or observation of participants activities (e.g. amount of water pumped, crop evapotranspiration). The eGROUNDWATER consortium will inform users on which data will be collected, how it will be processed and, in case the user or users to which each dataset refers could be identified, which measures will be taken by the consortium to guarantee the privacy of all users involved unless we obtain their explicit prior informed consent.

The project will comply with the international and national regulations in order to observe strict ethical standards and guarantee participants rights referring privacy protection:

- 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) (OJ L 119, 4.5.2016); and its corrigendum on 23rd may 2018.



- In January 2012, The European Commission proposes a comprehensive reform of the EU's 1995 data protection rules to strengthen online privacy rights and boost Europe's digital economy, 2012/0011 (COD).
- Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communication sector.
- Handbook on European data protection law by the European Union Agency for Fundamental Rights and the Council of Europe (2013).
- Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (Spain).
- Lei n.º 58/2019, de 8 de Agosto de 2019, Diário da República n.º 151/2019, Série I, Regulamento Geral de Proteção de Dados (Portugal).

5.1 Informed Consent and Assent

Informed consent and assent is required in when research involves the participation of users and is meant to guarantee the voluntary participation and understanding of the research purposes. Informed consent and assent consists of three components: adequate information, voluntariness and competence. Users should be clearly informed of the research goals, methods used for data collection and processing, possible adverse events, possibilities to refuse participation or withdraw from the research, at any time, and without consequences. Information to the users should be provided in such a way that they fully understand it and are aware of the implications of their assent.

eGROUNDWATER users involved will be provided at least with the following information beforehand through proper Information Sheets:

- The purposes of the research and information about what will happen with the results of the research.
- The experimental procedures and a detailed description of the involvement of the participants.
- All foreseeable risks related to the use of the measuring devices and interaction with the devices. We do not expect any risks in the participating in eGROUNDWATER.
- All benefits to the participants or to others which may reasonably be expected to occur.
- A description of the procedures adopted to guarantee the participants privacy.
- Contact details for researchers who can be contacted to answer pertinent questions about the research and the participant's rights.

DELIVERABLE D1.1: Data Management Plan

- A clear statement that the participation is voluntary, that the refusal to participate will involve no penalty and that the participant may decide at any time, to discontinue participation without penalty.
- General information about the research project.

eGROUNDWATER will demand and collect Informed Consent Forms from users concerning their understanding on the content of the Information Sheets and their consent to participate in the project. These forms will be safely collected, handled and preserved by the partners, and will be available to the PRIMA Foundation if required. UPV will have track of all the Informed Consent Forms collected by partners in each case study and will include the number of how many users signed the form in each case study in the Data Management Plan (D1.1).

5.2 Data privacy

Laid down as principles in the Charter of Fundamental Rights and the Treaty on the Functioning of the European Union, privacy and data protection are fundamental rights which need to be protected at all time. Data protection in eGROUNDWATER is meant to guarantee participants right to privacy and refers to the technical framework and security measures designed to guarantee that all personal data are safe from unforeseen, unintended or malevolent use. Users involved will be informed about the procedures used to collect, store and processing data.

The principle of proportionality will be observed, and no more data than necessary for the research and validation purpose will be collected. Personal data collection will be adequate and relevant. Data collected will be anonymous and, in case of point data collected by users or if they could be associated to a user's activity, the exact location of the measurements will not be publicly available unless explicit informed consent is given.

Storage, processing and communication of personal data will be undertaken in conformity with the EU General Data Protection Regulation (Regulation (EU) 2016/679 and its corrigendum on 23rd may 2018) and the relevant laws in each country. Each partner will identify the applicable local or national legal requirements in their countries and the competent authorities to provide the necessary authorizations for the handling of personal data. Data collected by eGROUNDWATER will be stored in a server of VisualNacert, which has a long-term experience on data collection and processing, complying will all legal requirements concerning data protection measures.

Any research action involving collecting personal data will be previously evaluated and approved by the Steering Committee, and no personal data will be gathered without the informed explicit consent of the participant. All personal information will be coded or anonymized as early as possible in the data processing.



6 eGROUNDWATER Data Record of collected and generated datasets

IDENTIFIER	VERSION	CASE STUDY	ORIGIN	VARIABLE	TYPE	FORMAT	SIZE	DATA OF UPLOAD	OPEN ACCESS	CONTACT FOR QUERIES	OBSERVATIONS
eGW_Coll_RqU_Trad_GwL_PTS_V1_0	1.0	Requena-Utiel	Traditional system (piezometer)	Groundwater level	Point time series	.txt	30 Kb	2020-07-01	Yes	hecmasor@upv.es	Data made openly accessible by the Jucar RBA (CHJ)