WATER4ALL DATA MANAGEMENT PLAN

Deliverable D3.1 - December 2022

<table>
<tr><td colspan="2" align="center">OUTPUT SUMMARY
Project information</td></tr>
<tr><td>Project Title</td><td>European Partnership Water4All – Water security for the planet</td></tr>
<tr><td>Project Acronym</td><td>Water4All</td></tr>
<tr><td>Call Identifier</td><td>Horizon-CL6-2021-Climate-01-02</td></tr>
<tr><td>Contract Number</td><td>101060874</td></tr>
<tr><td>Starting Date</td><td>1 June 2022</td></tr>
<tr><td>End Date</td><td>31 May 2029</td></tr>
<tr><td>Web-Site Address</td><td>www.water4all-partnership.eu</td></tr>
<tr><td>Coordinator</td><td>ANR</td></tr>
<tr><td>Management Team</td><td>Ariane BLUM, Juliette ARABI, Armelle MONTROSE, Claire TREIGNIER</td></tr>
<tr><td>E-Mail</td><td>Water4All@agencerecherche.fr</td></tr>
<tr><td>Phone Number</td><td>+33 1 78 09 81 20 / +33 1 73 54 81 43</td></tr>
</table>

<table>
<tr><td>Deliverable Title</td><td>D3.1 - Water4All Data Management Plan</td></tr>
<tr><td>Deliverable Number</td><td>D3.1</td></tr>
<tr><td>Work Package</td><td>WP3 - Pillar C: Science Policy – Policy – End-Users interface</td></tr>
<tr><td>WP leader</td><td>FORMAS / ISPRA</td></tr>
<tr><td>Nature</td><td>Report</td></tr>
<tr><td>Dissemination</td><td>Public</td></tr>
<tr><td>Editor (s)</td><td>Osman TIKANSAK (FORMAS)</td></tr>
<tr><td>E-Mail (s)</td><td>osman.tikansak@formas.se</td></tr>
<tr><td>Phone Number</td><td>-</td></tr>
<tr><td>Date of Delivery</td><td>19/12/2022</td></tr>
</table>

Verification by	Dominique DARMENDRAIL (BRGM)
Date	15/12/2022
Validation by	Claire TREIGNIER, Juliette ARABI (ANR)
Date	20/12/2022

Disclaimer: This document reflects the views only of the author, and the European Commission cannot be held responsible for any use which may be made of the information contained therein

TABLE OF CONTENTS

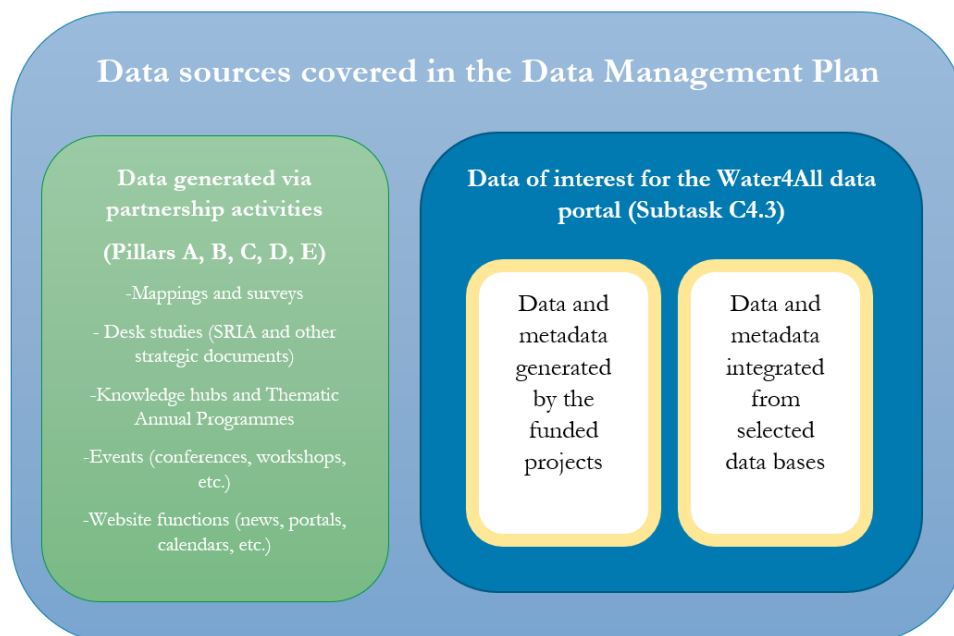# LIST OF ACRONYMS

ACRONYM	FULL TITLE
BRGM	French Geological Survey
DMP	Data Management Plan
DOI	Digital Object Identifier
DORA	Declaration on Research Assessment
EDB	Ethics and Deontology Board
EOSC	European Open Science Cloud
EC	European Commission
EU	European Union
FAIR principles	Findability, Accessibility, Interoperability, and Reuse
GDPR	General Data Protection Regulation
IPR	Intellectual Property Rights
ISO	International Standardisation Organisation
JTC	Joint Transnational Call
MST	Danish Environmental Portal
OD	Open Data
OGC	Open Geospatial Consortium
OI	Open Innovation
OS	Open Science
PCDE	Plan for Communication, Dissemination and Exploitation for Water4All Partnership
PCT	Partnership Coordination Team (led by ANR)
POPD	Protection of Personal Data
R&I	Research and innovation
RDA	Research Data Alliance
SRIA	Strategic Research and Innovation Agenda
Water4All	European Partnership on Water Security for the Planet
W3C	World Wide Web Organisation

SUMMARY

The Data Management Plan (DMP) of Water4All is prepared as a part of *Task C.4: Developing open science & open access to projects outputs and connected observatories and research infrastructures*. This DMP addresses two sources of data management in the Partnership (Figure 1).

The first source is related to the partnership's own activities. It covers issues related to personal and scientific data generated through implementation of additional activities under different pillars of the partnership (namely Pillars A, C, D and E), including surveys, mapping exercises, desk studies, knowledge hubs and web page functions among others. Such additional activities are essential to fulfil the expected impacts of the partnership from the European Commission's side by communicating, disseminating and exploiting the outcomes of Water4All activities in different formats, which frequently include various data sources.

The second source covers the data produced by the projects funded through the joint international calls (Pillar B) which might also use data from selected external databases. It is mostly quantitative and qualitative research data and metadata that fall under this category. Throughout this DMP it is made clear that all funded projects will be expected to produce their own data management plans to address specific issues related to their own research activities. One key task for Water4All will be to develop a data portal to compile the data produced by the funded projects and enhance the FAIR principles (Findability, Accessibility, Interoperability, and Reuse). In addition to compiling the data produced by the funded projects, the data portal will also engage with selected external data portals to support interoperability of existing water data across databases / portals. Design and development of Water4All data portal will be coordinated under *Subtask C4.3: Toolbox for managing water related data.* In this context, this DMP covers some of the preliminary issues concerning the development of Water4All data portal and some questions remain to be addressed as progress being made in the data portal development process.

FIGURE 1 - DATA SOURCES COVERED IN THE DATA MANAGEMENT PLAN

1. Introduction

The Water4All Partnership aims at enabling water security for all on the long term through boosting systemic transformations and changes across the entire research – water innovation pipeline, fostering the matchmaking between problem owners and solution providers. It gathers over 80 partners, R&I funders, environment ministries, local authorities, European, national and regional-scale networks, research performing organisations. Water4All proposes a portfolio of multinational, cross-sectoral activities, targeting a variety of actors, intending among others to produce, share and better communicate water-related knowledge and data, from local to global scales.

Delivering sound knowledge on water being the first specific objective of Water4All, the sharing of knowledge and data is of utmost importance for reaching the objectives of the partnership. Water4All will promote the application of open science, open data practices and the respect of the FAIR - Findability, Accessibility, Interoperability, and Reuse - principles across all its activities and pillars, while considering the European Union (EU) principles "as Open as Possible, as Closed as Necessary" (Opting-out option could be required in some cases requiring protection of intellectual property rights, or cases of national security and protection of public health). These principles should apply to data (or any digital object), metadata (information about that digital object), models and infrastructure newly developed.

2. Main principles of data management

The projects funded under Water4All calls should apply the FAIR principles through development and efficient implementation of Data Management Plans (DMP), which should be compliant with this Partnership DMP. At this stage, there is not a single DMP template that is fixed and applicable for all projects, but rather there are several adaptable tools that can be used for varying needs of the funded project while be compliant with this Partnership DMP. There is a list of free online DMP templates that are widely used and that will be recommended at a later stage by the Water4All partners at the end of this document (Appendix 1: Online tools for developing DMPs). In the upcoming months, Water4All will also propose a template for preparing DMPs in order to guide all funded projects in the achievement of this task by May 2023 (MS47-OS/OD/OS Policy for Water4All).

Other actions of the Water4All Partnership will also generate data which should be FAIR (e.g. surveys results for SRIA development, results of Water4All calls, impact assessment data, scoping activities, mapping studies, etc.). As a general principle, the data generated by these additional activities of the Water4All Partnership will be communicated, disseminated and exploited via communication channels and tools of the Partnership, including the Water4All website, online platforms, social networks and public events. Management of data that is generated by such additional activities will be subject to General Data Protection Regulation (GDPR) as usual and personal data will be made accessible only after informed consent of the person or via proper anonymization that leaves no room for identification of person. The list of communication activities to promote FAIR principles in Water4All activities are explained further in the "Plan for Communication, Dissemination and Exploitation for Water4All Partnership" (PCDE-Deliverable D1.2).

The concrete sharing of water-related data will culminate with subtask C.4.3, where interoperable Open Access / Open Data / Open Innovation interfaces will be developed for supporting data storage, persistence and access to products and data delivered by the Water4All activities, connecting various data sources from a single access, while providing tools for analysis and interpretation of the concerned data. It is expected that such a platform makes a significant contribution to the decision-making on water, by significantly reducing the effort for accessing a large amount of data. Following, the role of this Water4All platform is to be promoted,

so as to enlarge the amount of data connected to it as well as increase the number of its users, and therefore its impact.

This DMP does not deal with protection of personal data that is generated via additional activities of the Water4All Partnership, including all kind of communication activities and public events. Legal aspects pertaining to protection of personal data are covered under Article 15 of the Grant Agreement (Project 101060874) between the European Research Executive Agency (REA) and the beneficiaries of the Water4All. Measures for protection of personal data will be covered more in detail in the PCDE (deliverable D1.2), whereas the Partnership Coordination Team with the support of the Ethics and Deontology Board (EDB) will take responsibility on the GDPR implementation and monitoring.

Data typology: The data produced in Water4All Partnership can be split into two groups as the data coming from funded projects and the data generated by pillars' additional activities.

The first group of data is generated by funded projects and will deal with different (qualitative and quantitative) data formats depending on the scope of the calls. As an example, the 1st Joint Transnational Call (JTC) of the Water4All Partnership (2022) focuses on management of water resources under hydroclimatic extreme events, which inhabits different types of data concerning water scarcity, droughts, floods, water quality, groundwater and surface water monitoring among others. Given the expected variation of scientific areas and associated data formats over the years, it is important to develop a highly adaptable repository for data management with interoperable data interfaces.

The second group is broader in nature as it refers to all sorts of data generated as result of pillar activities including SRIA development, call development and open peer review of proposals, citizen science activities, training programmes, foresight and mapping exercises. Categorization of data formats for this group is not possible in advance due to the variations of these activities.

It is high likely that the call funded projects will use data generated by other projects and existing national/regional/global databases as necessary. Different sorts of external data will also be used for the implementation of additional activities of the Partnership.

Data origins: The data of the projects funded under the JTCs will originate from observation, experiments, theory and other sources, while various data collection methods will be used. Projects that are dealing with monitoring activities (water quality and quantity, hydrological cycle trends, water levels in aquifers, etc.) will rely on observational data over a period of time or in a continuous way for the projects that utilize real time data. In other projects, where lab work holds a central role (product development, chemical reactions, etc) experimental approaches may be prevalent. Models and simulations will be the focal point for projects that aim at developing decision support and scenario tools, thus the data will be derived from long term projections. Last but not least, some projects may utilize participatory citizen-science approaches to obtain the necessary data.

The data coming from additional activities will originate from various sources depending on the nature of activity. These may include foresight and insight studies via literature reviews, consultation activities with expert groups, surveys and monitoring activities among others.

The expected amount of storage required for the Water4All Partnership is not possible to estimate at this stage (e.g., due to unknown number of projects to be funded under the six JTCs planned in Water4All).

The data outcomes of the funded projects and additional activities will be of interest to the broader Water4All community which includes R&I funders, relevant ministries, local authorities, regional networks and research performing organisations. General public and decision makers at different levels are by default also among the target groups of the projects' data.

3. FAIR Data

A. Making data findable, including provisions for metadata

The ambition is to make all data discoverable with metadata. The funded projects will be requested to develop a DMP and to submit metadata of all the data resources directly generated by the project. The projects will be further encouraged to use standard unique identification mechanisms to identify metadata, datasets, features (HTTP URIs, Digital Object Identifiers).

As an overall measure to ensure FAIRness in data management activities, inventory of global standards (Plan S, Dora Principles-Declaration of Research Assessment) will be investigated and endorsed as appropriate. Current trends and best practices of open repositories will be evaluated including European Open Science Cloud (EOSC), EU Data Spaces, Open Geospatial Consortium (OGC), INSPIRE Directive, Zenodo, OpenAIRE when developing the data platform and toolbox.

In addition, approaches from existing research Infrastructures (e.g., European Plate Observing System - EPOS, eLTER) and projects such as GeoERA and its European Geological Data Infrastructure (EGDI), among others will also be considered.

Internationally agreed upon interoperability Standards and Best Practices (OGC, International Standardisation Organisation - ISO, INSPIRE, World Wide Web Consortium - W3C, Research Data Alliance - RDA, …) will be the basis for making Water4All data FAIR. This will be complemented by a close look at findings from the EOSC, and EU Data Spaces dynamics among others when developing the data platform and toolbox.

Specific questions to be addressed in due course within Water4All:

- **The metadata standards** that will be used: Depending on data to be shared metadata standards will be defined accordingly (e.g., ISO 19115, DCAT_AP, GeoDCAT_AP, other)
- **Where will the metadata be stored**? The metadata will be stored in the data platform of Water4All. The pilot version of the platform will be launched at month 24 (May 2024).
- **Will the metadata use standardised/non-standard vocabularies?** Metadata records will make an intense use of standards vocabularies. Re-use of pre-existing ones (e.g., from ISO, INSPIRE or other) will be favoured as long as they respect Linked Data principles.
- **Will the metadata for the call funded project datasets be available free-of-charge?** Yes.
- **Will the metadata for the call funded project datasets be harvestable by third-party applications and researchers?** Yes, by applying current best practices regarding metadata sharing (e.g., OGC CSW / API-Records, SPARQL endpoints)
- **Will you use a naming convention for your data?** This discussion is yet to happen within the project. It is anticipated that the core discussion will be applying FAIR principles (e.g.: regarding data semantics) prior to applying naming conventions for data.
- **Will you provide clear version numbers for your data?** This will depend on how frequently data is updated by the funded projects and whether the granularity regarding data update is important for the domain covered by the projects or no. For example, if we have real-time updates as projects continue to provide new data, then we might as well have version numbers for different data sets.
- **Will you provide persistent identifiers for your data? Using persistent identifiers means you have the freedom to move the data to new storage without impacting your users.** The persistent identification for data needs to be discussed during the project. It is too early to tell which solution will be deployed (e.g., DOI vs Uniform Resource identifier – http URIs) but there is a strong willingness within the project to apply FAIR principles and especially those in relations to unique identification and Linked Data Best Practices.

- **What services will you use to provide searchable metadata?** The software solution to store metadata for the project is yet to be identified but the foreseen solutions will have to support interoperable APIs (e.g., OGC CSW, OGC API – Records). OAI-PMH is also an option in discussion.
- **Which standardised data formats do you plan to use (such as PDF, HDF5, zip etc)?** Data format will highly depend on the types of data being produced. When serialization of data in a non-binary format is possible and makes sense, data formats from either JSON (e.g., JSON, GeoJSON, JSON-LD) or XML (e.g., GML) 'family' will be favoured. When data serializations in such formats will not be a reasonable solution, potential pre-existing binary formats well supported by the community producing this data will be reused.
- **Are the file formats you will use open?** Water4All ideal target is to only use open file formats (see response above regarding standardised data formats).
- **Do supported open-source tools exist for accessing the data?** Water4All data platform architecture will be based on semantics and APIs arising from the well-established data interoperability community (e.g., OGC, INSPIRE, RDA, W3C) and open-source implementations of those will be preferred. As a result, this will favour the use of open-source tools to access the data. When no such tool exist, effort will be made to contribute to the relevant open-source community to reduce the entry barriers.
- **Will you provide metadata describing the quality of the data**? As explained above, interoperable data semantics will depend on the data type being collected. When feasible data quality elements will be made available within the data itself. Otherwise, foreseen standards for metadata (see above) do also allow to document such information.

These questions will be discussed within Pillar C, in connection with Pillar B (*Developing new knowledge and innovative solutions for a systemic and inclusive approach of water challenges at operational scale*) and finally, with the Governing Board of the Water4All Partnership if decision is required.

B. Making data openly accessible

Accessibility of the call funded projects' data

The ultimate goal will be to make the data openly accessible at utmost level, by complying with the "as Open as Possible, as Closed as Necessary" principle. There might be ethical and legal issues that impact data sharing, which will be handled at project level and reflected accordingly in the project DMPs. Only the data that is considered accessible by the funded project consortium will be included in the Water4All data platform.

As a general condition of Horizon Europe, all projects funded under the Joint Transnational Call 2022 will support **Open Access to publications (Green Open Access** immediate or delayed open access provided through self-archiving, or **Gold Open Access** immediate open access provided by a publisher for the most important outputs). In the case of Green Open Access, partners will (as soon as possible and at the latest on publication) deposit a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications.

In addition, funded projects' partners will ensure open access to the deposited publication and bibliographic metadata. Information should be provided about all the other scholarly objects, tools and instruments that are needed to validate the conclusions of the publication. Specific rules will be provided to the funded projects in order to allow the Partnership to follow its impacts on Open Science principles.

In case of any **Intellectual Property Rights (IPR)** issue for some project data and products the consortium may decide to go for **Opt Out Option** and cover this within the project consortium agreement. In case of any IPR issue, projects can go for Opt Out Option before or after the signature of the consortium agreement (thereby freeing themselves from the associated obligations) on the following grounds:

- Incompatibility with the need for confidentiality in connection with security issues;

- Incompatibility with rules on protecting personal data;

- If there are other legitimate reasons not to provide open access to research data.

Depending on the funding rules of the Funding Performing Organisations, the costs related to the implementation of these provisions are eligible for reimbursement during the duration of the grant. In case of publications, the eligibility of such costs depends on the condition that the article should be published in a full open access venue. Publication fees in hybrid venues are not reimbursable. The European Commission's open access publishing platform, Open Research Europe provides a no-costs option for the projects to publish freely.

How will the data be made available?

The target is to make projects' data available through Water4All data platform.

The coordination of data publication will have to be jointly discussed with the funded projects.

Indeed, the overall target being to apply FAIR principles it won't help people interested in reusing data generated within Water4All if they are exposed differently regarding FAIR principles from a project web site and the data platform

Once the Water4All's data platform is launched in 2024 at pilot level, funded projects' data will be integrated to the platform gradually. The platform repository will be sufficiently secure and will ensure the data is backed up and can be recovered.

To increase accessibility explanatory guidelines and manuals will be available on the data platform to help users easily access the data.

Accessibility of the Water4All additional activities' data

Data and its associated products (reports, audio-visual material, meeting minutes, public deliverables, etc) will be made available in the Water4All's official webpage. To increase their publicity and accessibility, news about openly accessible products will be promoted via various channels such as newsletters, social media and online/physical public events. Details of such activities are explained further in the PCDE of Water4All. Furthermore, disclosure and accessibility of each public deliverable will be the responsibility of the partner who is in charge within the task. This means that the partner in charge should ensure proper communication with the Partnership Coordination Team (PCT) to make sure that the data and knowledge product is disseminated effectively.

C. Making data interoperable

Will you use a standard vocabulary for your data types to enable interoperability between data repositories?

Ensuring interoperability across different data types is a complex issue but will hold a top priority when designing the Water4All data platform. The funded projects will be requested to describe how they plan to make data will be interoperable in their DMPs. The topic will be discussed further:

- Throughout each project with a close interaction with the Water4All Data Platform Team (Task C4.3),
- And during the Water4All C4.3 workshops with partners to collect needs and contributions and test proposed semantics, tools and services.

Will you provide any support for data reuse?

As presented above, data will be described (semantic interoperability) and made accessible (technical interoperability) according to internationally agreed upon practices (OGC, INSPIRE, W3C, RDA …). This element is Water4All first building block for ensuring a better data reuse.

Guidelines will be made available to help researchers to better use the data. These guidelines can be derived for the projects themselves or new guidelines can be prepared for different data types if necessary. The staff behind the development and coordination of the data platform can also provide technical support (FQAs, support/info email, etc) to the platform users.

How long do you intend to support data reuse?

Data condensed in the platform will be accessible for the duration of Water4All Grant Agreement (2029). Possibilities to support the data platform for longer period of time will be investigated during the course of the Partnership.

D. Increase data reuse

The Internationally recognised open licences will be investigated and recommended to the funded projects to increase the reusability of data. Timeline for making the project data available for reuse will be set up by the funded projects in DMPs. As stated in the Annex 5 of the Grant Agreement (Specific Rules):

"Beneficiaries which have received funding under the grant must up to four years after the end of the action use their best efforts to exploit their results directly or to have them exploited indirectly by another entity, in particular through transfer or licensing

If, despite a beneficiary's best efforts, the results are not exploited within one year after the end of the action, the beneficiaries must (unless otherwise agreed in writing with the granting authority use the Horizon Results Platform to find interested parties to exploit the results."

The Water4All Data Platform, on the other hand, will be launched in May 2024 and gradually populated with data retrieved from the projects. Depending on the availability of the data sources, the economic model for keeping operational Water4All data platform after the end of the Partnership will be assessed during the Partnership (e.g., via a European Research Infrastructure).

In order to achieve quality assurance of data in funded projects, data quality procedures will be selected by the projects data teams and documented in their deliverables. When these data are submitted to the Water4All Data Platform, there will be a check-list to make sure that all mandatory meta-data fields are filled in.

4. Allocation of resources

Resources for managing research data: When developing the Water4All database, relevant existing capacities will be used to utilize the resources more effectively. For example, the infrastructures of the French Geological Survey (BRGM) and Danish Environmental Portal (MST) will be used as baseline to build up the Water4All data platform. Funded projects might also use national infrastructures and databases, which should be clarified in their project DMPs. Other national data repositories that will be interfaced with the Water4All platform will be identified before the pilot launch of the data platform. During the first phase of the Grant Agreement, 93.6 PM will be allocated by the partners in Task C4.3 to the activities related to developing a data platform and related toolbox.

Resources for ensuring open accessibility of data: For the funded projects, the costs for publishing research outcomes in open access peer reviewed journals will be eligible for reimbursement. The EC Top-Up will be the source for reimbursing this cost.

Resources to ensure data reuse after the Partnership ends: All options regarding resources to ensure continuity of the data platform in an archive after the lifespan of Water4All will be investigated during the foreseen consultations and workshops.

Resources to store projects and personal data gathered through JTC proposal submission platform: In-kind eligible costs will be reimbursed by the funding agency who is hosting the submission platform services.

Resources to store other data related to the additional activities in Water4All website, intranet and document sharing platform (Microsoft TEAMS): These costs will be covered in-kind by the PCT and reimbursable costs will be reported to EC under the coordination and management costs.

5. Data security

Overall data security of the Partnership will be subject to EU, international and national laws on data protection. Protection of Personal Data (POPD) & Non-EU Countries framework will be developed with the support of the EDB to demonstrate the engagement of all Water4All partners to ensure compliance with the European regulations. Both for the funded projects and for the additional activities of the Partnership each party will remain responsible to set up internal processes to comply with the personal data regulations.

The technical specifications for data security of the Water4All Data Platform will be addressed in due course as the process evolves. Funded projects on the other hand will handle data security of their research individually and describe the details in their DMPs. The privacy policy of the Partnership will include further information about how the personal data will be used.

6. Ethical aspects

Any relevant ethical aspects should be dealt with due diligence by the funded projects in their research plans. It is expected that the by the time that projects produce their data the ethical aspects should not be a concern anymore, so that data can be integrated into the data platform directly.

Ethical aspects concerning the additional activities of the Partnership will be subject to the provisions of the Consortium Agreement, and thus binding regulations at EU and national levels.

7. Other

Both the projects and Partnership data will not be subject to any additional data management procedures for the moment. This DMP is considered as a living document and will be updated regularly, so the future iterations may include new provisions.

APPENDIXES (to be developed in the next version):

- APPENDIX I: ONLINE TOOLS FOR DEVELOPING DMP
- APPENDIX II: LINKS TO OPEN DATA/OPEN ACCESS POLICIES OF WATER4ALL PARTNERS

water4all@agencerecherche.fr
www.water4all-partnership.eu

Grant Agreement n° 101060874