PROJECT DELIVERABLE REPORT

NAIADES

Greening the economy in line with the sustainable development goals

D1.7 DATA MANAGEMENT PLAN

A holistic water ecosystem for digitisation of urban water sector

SC5-11-2018

Digital solutions for water: linking the physical and digital world for water solutions

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 820985"

Document Information

Grant Agreement Number	820985	Acronym	NAIADES
Full Title	A holistic water ecosystem for digitization of urban water sector		
Topic	SC5-11-2018: Digital solutions for water: linking the physical and digital world for water solutions		
Funding scheme	IA - Innovation action		
Start Date	1stJUNE 2019	**Duration**	36 months
Project URL	www.naiades-project.eu		
EU Project Officer	Alexandre VACHER		
Project Coordinator	CENTER FOR RESEARCH AND TECHNOLOGY HELLAS - CERTH		
Deliverable	D1.7 – Data Management Plan		
Work Package	WP1 – Project Management, Quality Assurance and Reporting		
Date of Delivery	**Contractual**	M6	**Actual**
Nature	ORDP: Open Research Data Pilot	**Dissemination Level**	PU-PUBLIC
Lead Beneficiary	CERTH		
Responsible Author	Dionysis Bochtis	Email	d.bochtis@certh.gr
		Phone	
Reviewer(s):	Leonardo Alfonso (IHE)		
Keywords			

Revision History

Version	Date	Responsible	Description/Remarks/Reason for changes
0.1	05/11/2019	CERTH	Report write-up (A. Anagnostis)
0.2	11/11/2019	ALL	Inclusion of partners' contributions
0.3	12/11/2019	CERTH	Internal Review (G. Banias, G. Tsiotra)
0.4	19/11/2019	IHE	Leonardo Alfonso's review and comments
0.5	22/11/2019	VUB	Vagelis Papakonstantinou's review and comments
1.0	…	CERTH	Review and Release
1.01	14/07/2021	CERTH	Revisions addressing PO's comments after RP1
1.1	21/07/2021	IHE, AIMEN, VUB	Review and comments
1.2	30/07/2021	CERTH	Final review and corrections
2.0	30/07/2021	CERTH	Review and Release

Contents

Abbreviations

DMP	Data Management Plan
OA	Open Access
IPR	Intellectual Property Rights
ORD	Open Research Data
TRL	Technology Readiness Level
DMPO	Data Management Portal
API	Application Programming Interface

AI	Artificial Intelligence
WTP	Water Treatment Plant
HTTPS	Hypertext Transfer Protocol Secure
DCA	Data Collection Aggregation
dWTP	Drinking Water Treatment Plant
FTP	File Transfer Protocol
ICT	Information and Communications Technology
IoT	Internet of Things
ETSI	European Telecommunications Standards Institute

RP1 review comments and responses

PO's comments		Response
Shortcomings	NAIADES has been affected by data lacking because of confidentiality issues that the data management plan hasn't managed.	Since this Deliverable was submitted in M6, the project's progress made until then was at a preliminary stage. Therefore, some data access issues could not be forecasted at that time.
Recommendations	Update the data management plan addressing the data confidentiality issues that popped up so far.	Data confidentiality issues have been addressed in the duration of the project, and as it has been mentioned thoroughly within the revised deliverables, data-related issues, such as the lack of data, or the WTP-related data, had actually no tangible negative impact on the project's progress. Actions required for addressing lack of data issues due to confidentiality are presented in Section 9. Moreover, in order to gain a clearer overview about the data generated in the first period of the project, a questionnaire has been sent to partners. This questionnaire is based on Horizon 2020 FAIR Data Management Plan (DMP) template[1]. Partners' inputs are presented in Annex I. According to those inputs, data management plan has been updated and the main outcomes of this table are highlighted in Section 9.

[1] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

1 Summary

This report describes the Data Management Plan (DMP) of the NAIADES project and focuses on the recommendations and guidelines to the project's partners on how to collect, generate, manage and re-use the NAIADES data. More specifically, this document is a report that specifies how data and research publications will be collected, processed, monitored, catalogued, and disseminated during the project lifetime, as well as after the end of the project.

2 Introduction

European-funded projects such as NAIADES usually produce large sets of data during their lifetime. Depending on the discipline, the data could derive from various sources such as laboratory testing, field trials, social science research and various observations. One of the major problems in these kinds of projects is the uncertainty of what will happen to the data after they are analysed and the project has finished. In fact, the majority of data created can be of high value for other researchers, but because they are either stored on local servers and/or miss crucial metadata, their potential value is lost. Thus, in line with the purpose of this document, all consortium members need to take into account a plan about the data that they will produce during their work.

The main goal of the DMP is to form an analysis of the basic elements of the data management policy that will be used within the NAIADES project by the consortium members with regard to the project data. The DMP covers the types of research data that will be generated or collected during the project, the standards to be used, the way the data will be preserved and what parts of the datasets will be shared for verification or reuse.

The nature of the DMP is continuous, like a living document, which will evolve within the duration of the project, particularly whenever significant changes arise such as dataset updates or changes in Consortium policies. This document is the first version of the DMP, delivered in Month 6 of the project. It includes an overview of the datasets that will be generated/acquired within the project, and the specific conditions that govern them. Although this report already covers a broad range of aspects related to the NAIADES data management, the upcoming updates will offer more details on particular issues such as data interoperability and practical data management procedures implemented within the NAIADES project consortium.

In the following sections, the DMP provides an overview of the datasets, which will be generated/acquired/handled in the NAIADES project. In detail, the origins and formats of all data are described, as well as their connection to their particular WPs. Furthermore, the purpose of the data collection as well as information on the use of the collected/handled data is highlighted within the DMP. The data that will and will not be made openly accessible, including detailed justifications of these decisions, are also clearly presented. Moreover, details on the data repositories and management platforms where the data will be stored are described. The final version of this deliverable will provide a detailed report on the whole data management strategy followed by NAIADES, and will be continuously improved based on open issues and questions that will be addressed as the project progresses.

3 Data summary

There are several types of data that will be created/gathered/acquired within the project's duration. Most of the partners will be handling data in one way or another, therefore a concrete data management plan would offer clarity, transparency and a proper documentation to all the partners that are involved. Table 1 shows an initial overview of the type of data, the possible formats, their sources and description.

Table 1. Data overview and description.

#	Type	Format	Origin	Uses	Description and Purpose
1	Quantitative numerical datasets	.xlsx, .csv, .json, .xml	Public and confidential	NAIADES' research activities rely on mathematical and arithmetic analyses. Quantitative variables will provide the necessary information for the proper scientific and technical research of this project.	Numerical data will be describing measurable quantities that are associated with the operation of the use cases and their processes. Algorithms and models will be built based on such data, and new data will be produced as an outcome of those models.
2	Categorical data	.json, xlsx, .csv, .tsv	Public and confidential	Categorical data will be used in tandem with the quantitative data and help build robust models and forecasts.	Categorical data describes variables that can be summed up in categories or variables that can be quantified in order to be used in arithmetic models.
3	Qualitative data	.doc, .pdf	Public and confidential	Qualitative data will offer insight in the NAIADES' services by improving models and approaches with human knowledge and input.	Qualitative data are data based on abstract concepts that ultimately need to be translated into numerical data in order to be valuable for the NAIADES' services.
4	Water distribution simulation (EPANET)	.dxf, .ascii, .inp	Confidential	They will be used to model water distribution systems. Their use will be to build a model for understanding the movement and fate of drinking water constituents within distribution systems.	Long-period simulation of the hydraulic and water quality behavior within pressurized pipe networks, which consist of pipes, nodes (junctions), pumps, valves, storage tanks, and reservoirs, track the flow of water in each pipe, the pressure at each node, the height of the water in each tank, a chemical concentration, the age of the water, and source tracing throughout the network during a simulation period.

5	Geographical data (GIS)	GIS vector, GIS raster	Public	Shapefile data will be used for storing geometric location and associated attribute information.	Shapefile files are geospatial vector data for geographic information system (GIS) software.
6	Written and published material	.pdf, .ppt, .docx	Public and confidential	Written and printed material will be produced in cases such as conferences, meetings and other public events. Additionally, publications and deliverables will be publicly available (except for confidential deliverables).	Brochures, newsletters, project reports, journal/magazine articles, conference, presentations/posters, scientific publications, public deliverables for dissemination of the project.

4 FAIR data

FAIR data is EU's approach for Findable, Accessible, Interoperable and Reusable data that will boost the dissemination of information and the easy exchange of data within the state member across Europe. This section describes how the NAIADES project envisions to make its data FAIR.

4.1 Making data Findable, including provisions for metadata

The data produced/collected within the NAIADES project will be discoverable and will have metadata that follow a standard identification mechanism. The data and management platform is presented later on this document. The naming conventions have been decided by all members of the consortium and can be found in D1.1 "Project Handbook" that was submitted on month 3 of the project's duration. The "NAIADES" keyword will be present in all files so that the data can be easily discoverable and correlated to the project. Versioning in a standard procedure as well within the project in order to keep track of the work done. Metadata will be created appropriately to each type of data that is deemed necessary, which will be created/used within the project. The type of metadata will be left on each partner that will be responsible of creating/using the data for their work.

4.2 Making data openly Accessible

Data accessibility is a priority within the NAIADES project. Except the cases where there is a proper justification, all data will be publicly accessible. In the cases where confidentiality of data is required, an alternative solution will be provided. This information is given in **Error! Reference source not found.**.

Table 2. Data accessibility.

#	Type	Availability	Justification	Alternative
1	Services portfolio	Public	A document listing the offered services will be open to everyone to attract potential partners and/or clients.	
2	Brochures, flyers, etc.	Public	All printed material will be used for the dissemination of the project, so that it will receive recognition and attract interest.	
3	Weather forecast	Confidential	The weather forecasting service will be provided to the NAIADES services and its users.	Several services provide forecasts (OpenWeatherMap,

				WeatherUnderground, etc.)
4	Water demand	Confidential	The water demand forecasting service will be provided to the NAIADES services and its users.	Not yet defined.
5	Water flow	Confidential	The water flow monitoring service will be provided to the NAIADES services and its users.	Not yet defined.
6	Water consumption	Confidential	The water consumption forecasting service will be provided to the NAIADES services and its users.	Not yet defined.
7	Water pressure	Confidential	The water pressure monitoring service will be provided to the NAIADES services and its users.	Not yet defined.
8	Water quality	Confidential	The water quality monitoring service will be provided to the NAIADES services and its users.	Not yet defined.

The details of the publicly available data mentioned in Table 2 are presented in **Error! Reference source not found.**.

Table 3. Details of accessible data.

#	Type	Location	Level of accessibility	Type of availability and required software tools	Information on metadata and additional data information
1	Services portfolio	NAIADES Services Platform	Public	Web-browser for accessing the link and a PDF-viewer (or a browser) to view the file	No metadata or information required
2	Brochures, flyers, etc.	Partners' premises, conferences, workshops, etc.	Public	Web-browser for accessing the link and a PDF-viewer (or a browser) to view the file	No metadata or information required

4.3 Making data Interoperable

Standard vocabularies will be used for all data types present in the project. A detailed depiction of these vocabularies is presented in D9.10 "Report on standards (a) used and (b) generated in NAIADES" submitted in month 3. In case there is an uncommon vocabulary, a clear mapping will be provided in order to facilitate its use. Thus, the project's data will be interoperable and easy for sharing among researchers, institutions and organizations.

4.4 Increase data Re-use (through clarifying licences)

The reusability of data will be based on common licenses (e.g. Creative Commons). However, some of the

data might need to be held as confidential for an amount of time for specific reasons (e.g. patents). The data will be stored in Freedcamp, the selected data repository which is described in a separate section below, for all the duration of the project, plus some time after the conclusion of the project, which will be decided later on by the consortium.

There is also a plan to use Zenodo for sharing documents and datasets openly with the community outside the project. The Zenodo repository is generally recommended due to its support by the EU commission under OpenAire.

5 Allocation of resources

The consortium will use the Freedcamp management platform, which is described in detail in Section 6.2 as a repository for data safekeeping, accessible. All the deliverables containing accessible as well as confidential data, will be uploaded and stored there. Freedcamp's security mechanisms will ensure that the data is safely stored for a long period even after the completion of the project.

The handling of Freedcamp as a repository, as a general management platform and as all data management is preferred in the first stage of the project with the possibility of adapting also Zenodo. The management of Freedcamp related to the project inside NAIADES, is under the responsibility of the coordinator.

All publications related to the research conducted during the project, will be submitted to scientific journals that comply with an open-access policy, and the fees will be covered by the budget allocated by the H2020 grant to the NAIADES project.

6 Data security

6.1 IPR and copyrights

NAIADES project covers some high-TRL technologies that aim to develop marketable solutions. The project consortium includes partners from the private sector, end users and demonstrators. Each of the consortium members is qualified for Intellectual Property Rights (IPR) on their technologies and data, on which their economic sustainability could be at stake. Thus, the NAIADES consortium will protect that data and get approval of concerned partners before every data publication.

The data management portal, i.e. Freedcamp, will be equipped with authentication mechanisms to handle the identification of the persons/organizations that download the data, as well as the purpose and the use of the downloaded dataset.

6.2 Data management portal

The Data Management Portal DMPO, a web based portal namely Freedcamp, is being used within the NAIADES project for the purposes of the management of the various datasets that will be produced by the project, as well as, for supporting the exploitation perspectives for each of those datasets. Freedcamp Portal is flexible in terms of the parts of datasets that are made publicly available. Special attention will be given to ensure that the data made publicly available violate neither IPR issues related to the project partners, nor the regulations and good practices around personal data protection.

Access to The Freedcamp Portal is given through a web-based platform, which enables its users to easily access and effectively manage the various datasets created throughout the development of the project. Regarding the user authentication, as well as the respective permissions and access rights, the following three user categories are foreseen:

- **Admin:** the Admin has access to all of the datasets and the functionalities offered by the DMP and is able

to determine and adjust the editing/access rights of the registered members and users (OA area). Finally, the Admin is able to access and extract the analytics, concerning the visitors of the portal.

- **Member:** when someone successfully registers to the portal and is given access permission by the Admin, she/he is then considered as a "registered Member". All the registered members will have access to and be able to manage most of the collected datasets. Knowledge sharing and public documents, apart from the admin and the registered members, as Open Access (OA) area will be available for users who will not need to register and they will have access to some specific datasets, as well as to project outcomes. Figure 1 shows the home screen of Freedcamp.

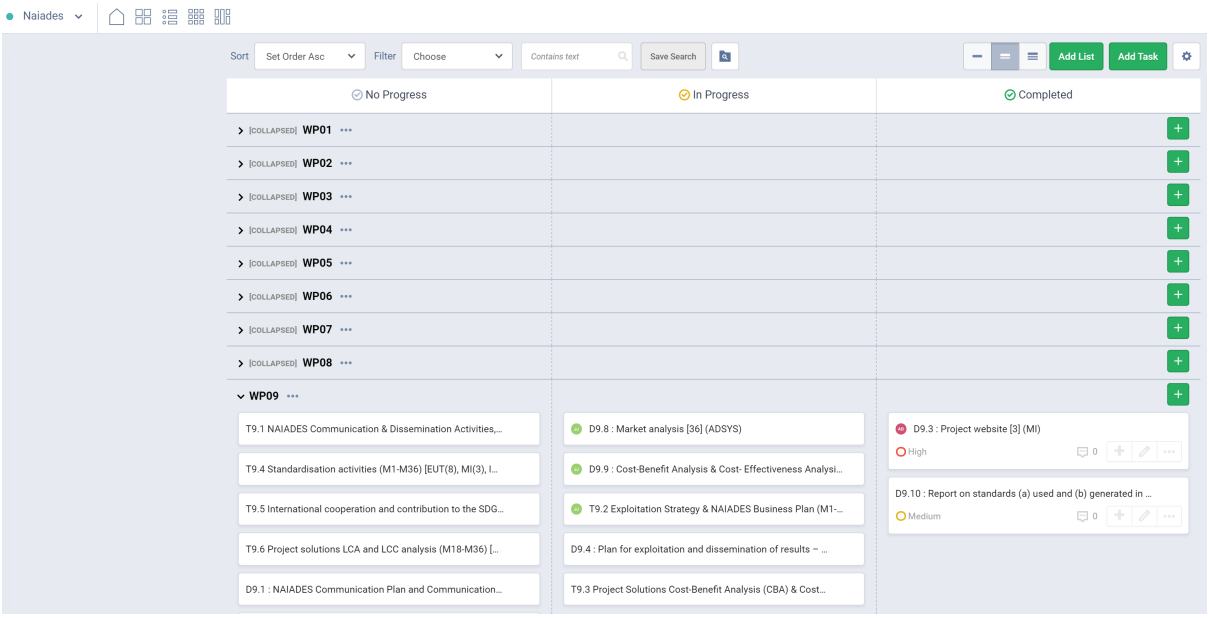

Figure 1. Freedcamp home screen.

Freedcamp portal is easily and effectively managed by the members. A variety of graphs, pie charts etc. are going to be employed for helping members to easily understand and elaborate the data. In particular, the architecture of the portal presents special interfaces organized to comply the information. All tasks and datasets available in the DMP are accompanied by a short description of the item (Figure 2).

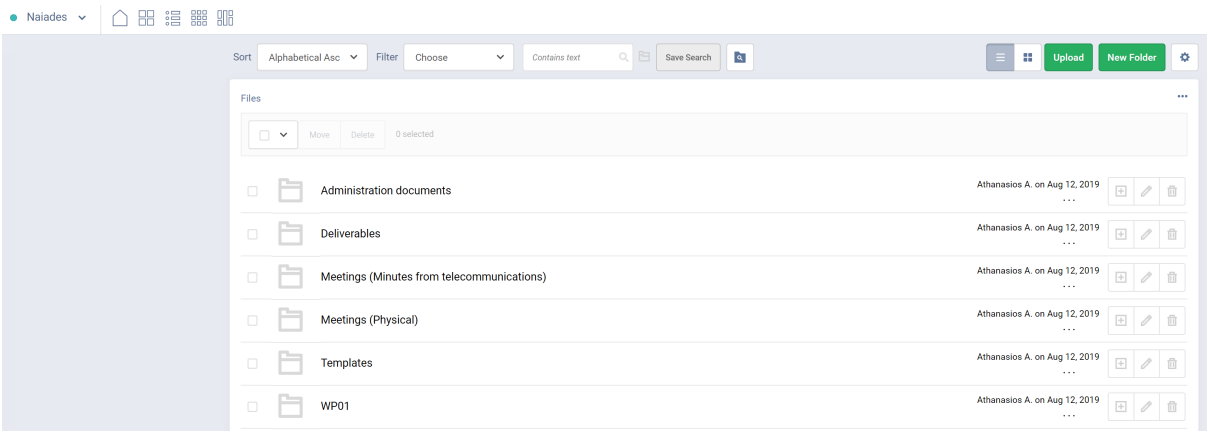

Figure 2. Structure of data file system.

Datasets are structured in three different folders into Freedcamp portal; Tasks, Discussion, Files. Draft documents and deliverables, and other data are uploaded on specific tasks folders; final version documents will be uploaded into the file section of the appropriate folder. In addition, technical and progress meetings are being scheduled in the Freedcamp portal calendar (Figure 3).

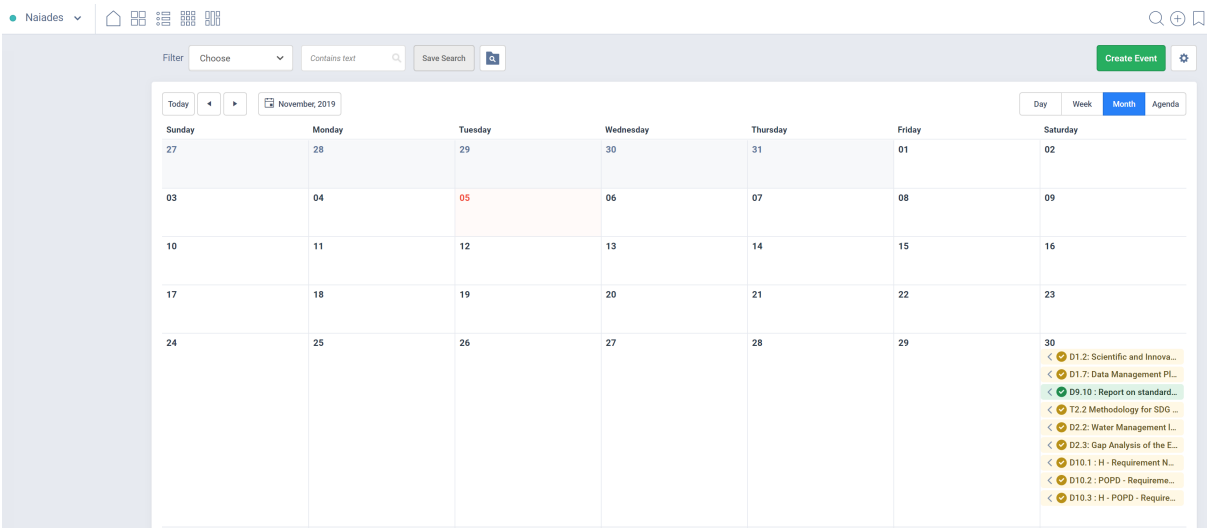

Figure 3. Calendar functionality with impending deadlines.

7 Further support developing the DMP

The DMP is considered as a constant "work-in-process" and will be continuously updated throughout the duration of the project. Therefore, some issues will need to be addressed at later stages as work progresses. An indicative list of issues that might come up is given in Table 4.

Table 4. Issues for future resolve regarding the DMP.

Category	Potential issues
Data interoperability	- What is the strategy for ensuring that the data produced in the project are interoperable? - How can make sure that data exchange and reuse between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, compliant with available (mostly open) software applications, and in particular facilitating re-combinations with different datasets from different origins) is allowed? - What data and metadata vocabularies, standards or methodologies will be followed for data interoperability? - Will standard vocabularies for all data types will be used, to allow inter-disciplinary interoperability?
Data re-use (through clarifying licences)	- How long will the data be available for reuse? - How will the data be licensed so that they can be reusable? - How and when will the data be available for reuse? - Are the data produced/used by third parties and will this affect their availability for reuse? - What are the data quality assurances?
Allocation of resources	- What is the long-term strategy for preservation of data, taking into account cost for storage and value of data?
Data security	- What measures are being taken for data security (sensitive data, data recovery, etc)?
Other issues	- Are some national/sectorial/departmental strategies for data management used? - Are the best practices and feedback mechanisms for the data management platform (Freedcamp) followed properly?

8 Conclusions

This report presents the DMP and describes the data information that will be generated during NAIADES project and the challenges and constraints that need to be taken into account for managing it. In addition, it describes the procedures and the infrastructure used in the project to efficiently manage the produced data, named as Freedcamp Management Portal. The DMP is identified as a starting point for the discussion with the partners in the beginning, and later on with a broader community, about the NAIADES data management strategy and reflects the procedures planned by the work packages at the beginning of the project.

An extended discussion has been conducted within the consortium partners, in order to get a general view on the kind of data they were expecting to produce and collect during the project. At the moment not all activities are planned to generate data that would potentially cover all EU's expectations, however, the DMP is a living document and it can be the case that this situation evolves during the lifespan of the project. Thus, the DMP will be updated and augmented with new datasets and results during the project lifetime.

Regarding storage information, documents generated during the project will be stored in Freedcamp Portal which is the all-around project management system of the NAIADES. This information, data and documents produced during the project will be protected for a period of time after the project completion, as it is described in GA.

9 Updates on DMP

One of the key factors for the development of the NAIADES AI ecosystem is the generation/collection/process of large amount of data. Partners responsible for data handling have planned to perform the aforementioned actions according to the NAIADES' Description of Action (DoA) and the designed architecture. The initial DMP presented in this document had been submitted at M6. At this point, the core architecture of the project had not been finalised yet. As a consequence, partners were not totally aware of the issues that may arise concerning the generation, access, collection, and process of data during their manifold tasks within NAIADES. After the first reporting period, one main issue that has been reported was the lack of data (e.g., WTP-related data) caused by data confidentiality. Limited accessibility to data in some tasks, due to confidentiality issues, has led to deviations from the initial plans. This, in turn, impeded the collection and access of raw data which is vital for data-driven modelling. Therefore, alternative actions are required in order to mitigate these contingencies. Since DMP is considered to be a living document, it is updated throughout the duration of the project in order to address some issues that may occur as work progresses.

When lack of data issues arise, partners should come up with counter-measures for the acquisition of relevant data. Based on the updated Project Handbook (D1.1), partners should report the risk of the counter-measures to the respective boards and the Scientific & Technical Manager, so that they can estimate the possibility of the reported risk and the impact it can bring upon the project. A first approach that partners should adopt in order to overcome lack of data issues is the acquisition of historical data from respective services of the use cases. Historical data retrieval can be done by exploiting other data sources, such as Application Programming Interfaces (APIs) containing data relevant to the use cases. In particular, open APIs (or public APIs) are publicly available APIs that enable developers to access web services and datasets. Their key features are:

- Developers are able to use them under relatively few restrictions (e.g., registration to the service),
- They typically provide open data which is freely available for everyone to use and republish (without any restrictions from copyright, patents or mechanisms of control), and
- They are designed based on open standards that specify what content is publicly available.

Nevertheless, there are some issues that need to be taken into account during the usage of APIs:

- Unauthorised changes to provided information,
- Information leakage, and
- Interference with legitimate activity.

Consequently, it should be evaluated if data sources are compliant with the General Data Protection Regulation (GDPR)[2] or any other national law, especially when personal data is retrieved. More specifically, partners should make sure that no personal data is collected at user level. In addition, when personal data is collected, it should be aggregated with anonymous data, if it is feasible, in order to minimise the possibility of data subjects' identification. These guidelines should also be adopted when data from social media platforms are going to be used. However, when acquisition of historical data from respective services of the use cases is not achievable, lack of data can also be addressed by conducting laboratory measurements or through simulations.

Furthermore, in order to get a better view on the processes under which data is generated/accessed/collected in the first reporting period, it is essential to gather details on how data is handled within NAIADES. Therefore, a questionnaire regarding various aspects of data processing has been circulated among the involved partners. This questionnaire is based on the Horizon 2020 FAIR Data

[2] *https://eur-lex.europa.eu/eli/reg/2016/679/oj*

Management Plan (DMP) template, according to the EU principles for Findable, Accessible, Interoperable, Re-usable data (FAIR principles). This document should not be considered as a technical implementation that strictly follows the FAIR principles; it is actually inspired by these principles as a general approach.

The answers that partners provided, in the context of this questionnaire, are presented in *Annex I*. The main outcomes of this table are summarized below:

- *Data is discoverable with metadata, identifiable and locatable through standard identification mechanisms*: Data includes metadata which provide information about the location and the collection time of the measurements as well as unique identifiers. As a result, data collected by each integrated sensor can be identified by IoT platform databases and services.
- *Measures taken for data security*: NAIADES technological components integrate authorisation and authentication mechanisms for data handling. Regarding the collection of raw data from Data Collection Aggregation (DCA) component, a security mechanism based on the widely used HTTPS protocol is utilised. In addition, data sent from various components to IoT platform are signed with Keyless Signature Infrastructure (KSI) signatures. Concerning FIWARE components, Keyrock and Wilma are exploited for authentication purposes.
- *Data naming conventions*: Schema.org[3] has been extensively utilised for the naming conventions of data. Schema.org vocabularies provide schemas for structured data. These vocabularies include entities, relationships between entities and actions and can be easily extended. In this way, shared vocabularies can be easily adopted and maximise the benefits of its usage.
- *Openly accessible data*: Apart from data generated by operations related to specific services and users (e.g., weather forecasting, water monitoring), all other data remains publicly available.
- *Data access*: The data is being made accessible through NAIADES modules (e.g., HMI-Human Machine Interface) and components (e.g., Marketplace). Moreover, data is accessible through a Water Standardization Catalogue[4] which facilitates the seamless access to various types of data.
- *Documentation on data access software*: In general, software about data access comes with an appropriate documentation. More specifically, the documentation about the software that provides access to data stored in the Water Standardization Catalogue is included on github[5]. Concerning NAIADES components (e.g., Marketplace) the APIs will be described in the documentation page of APIs' User Interface (UI).
- *Software inclusion for data access*: In most of the cases, software regarding data access is open-sourced (e.g., Water Standardization Catalogue) or freely available (e.g., EPANET[6]). Furthermore, details on how the relevant software can be used are elaborated on respective Deliverables.
- *Data access under restrictions*: When restrictions on data access are applied, access to data stored in the IoT repository is provided to users according to their role. In such cases, users' role is verified and the proper rights are duly provided. Concerning the data of DCA, DCA repository is private.
- *Interoperable data*: Since NAIADES interoperability is based on the FIWARE platform[7], an open-source platform for context data management, any other institution, organisation, or researcher from other projects compliant with this platform is able to exchange and re-use data generated by its components (e.g., Next Generation Services Interfaces- NGSIv2 and NGSI-LD). Furthermore, data related to the Water Standardization Catalogue can be exchanged and re-used, as it is an open-source project. Regarding part of the Marketplace data, it can be accessible by institutions.

[3] *https://schema.org/*
[4] *https://aolite.github.io/naiadesStandardization/#/*
[5] *https://github.com/aolite/naiadesStandardization*
[6] *https://www.epa.gov/water-research/epanet*
[7] *https://www.fiware.org/*

In conclusion, data confidentiality issues have been addressed amidst the project, and had no tangible negative impact on the project's progress. However, the data information that is being generated during NAIADES project will remain under continuous monitoring and DMP will be updated accordingly if other issues arise.

Annexes

Annex I: Partners' inputs on Horizon 2020 FAIR Data Management Plan (DMP) Questionnaire

Data Management in NAIADES							
Questions	***SIMAVI***	***Eurecat***	***UDGA***	***JSI***	***CUP Braila***	***AMAEM***	***AIMEN***
What is the purpose of the data collection/ generation and its relation to the objectives of the project?	SIMAVI collects raw data from CUP Braila's sensorial system and model it in order to send it to the IoT platform so that it is analysed and persisted in the data base. Based on the data, the AI services will make predictions and statistics to help CUP Braila have a better water management.	The information collected from our side refers to specific standards and interoperability strategies. It has served to elaborate a standardization strategy inside NAIADES. Moreover, this collected information has been stored into the online version as a JSON file.	The Carouge pilot data has been collected to UDGA's DCA in order to format to the common data models and provide it to the NAIADES platform where all data is collected and served for NAIADES services.	JSI is responsible for a delivery of various technical (mostly AI) solutions. The input JSI needs are collections of data (provided by other partners), while the output is, again, of form of data.	The data collected and sent by CUP Braila covers: flow sensors, pressure sensors, noise sensors and weather data. This data is used to calibrate and train the AI systems that are being developed within the project.	Data collection and generation is required for the use cases' applications (UCA1 - water demand forecast; UCA2 - sewage saline infiltration detection; UCA3 - municipal water consumption platform)	We need historical data of water quality parameters to train AI for water quality predictions. Also, real time data to predict future values. Related to this problem, we also need historical, real time and forecasted values of weather to include in the models. We also need historical and real-time data of water quality and treatments dosages applied in drinking water treatment plants (dWTP); for the same purposes as before: training of models, and predictions.
What types and formats of data will the project generate/collect?	As raw data, Data Collection Aggregation (DCA) collects CSV files with raw data from SCADA system of CUP Braila that contains attributes of type: date, numeric and string. DCA will model the CSV files and generate JSON files which contain the attributes of types as before mentioned that will be send to IoT platform.	JSON file containing the standards.	JSON and JSON-LD.	JSI generates data in JSON format based on FIWARE standards.	CUP Braila supplies/ has supplied the project with data covering pressure, noise, demand, weather, location and hydraulic models.	Network sector water volume (m^3), Conductivity (µSiemens), Waste water level (m), Waste water flow (m^3/s), Water consumption in public facilities (m^3), Green area extent (m^2), Number of users in public buildings (users)	All the data types we require are numerical, float numbers.
Will you re-use any existing data and how?	SIMAVI used some historical data provided to the partners in charge of AI services with the scope of improving/ training the algorithms.	The data is available here: https://aolite.github.io/naiades	NAIADES AI services are trained with historical data, and NAIADES	No.	CUP Braila does not compute the data within the project.	Selected data from AMAEM's systems (network sector meters and smart meters) will	For the Water Quality Prediction, we started using data online to create the first versions of the models. But we require, data from

		Standardization /#/.	platform provides a historical database to support it.			be reused. New sensors (conductivity, level and flow) will provide additional data for UCA2.	the real problem to improve the results. For the treatment suggestions solution, there were no data to re-use.
What is the origin of the data?	For the weather use case, DCA consumes data from Romanian National Meteorological Administration and for the water demand use case and leakages use case, DCA consumes data from hydraulic SCADA system of CUP Braila.	The data has been originated in D9.10 of NAIADES	Sensing data from the sensors installed in Carouge and Open data on Carouge weather.	Naiades platform and JSI algorithms.	CUP Braila is the source of the data sent, which originates from various sensors within the target area.	Selected data from AMAEM's systems (SCADA and smart meters) is pushed to an FTP.	For the water quality prediction, we started using data from https://gemstat.bafg.de/applications/public.html?publicuser=PublicUser#gemstat/Stations But the origin of the data of the final solution will be the new sensors installed in Carouge Fountains. And the weather data from Carouge local weather station (provided by UDGA) and results from the weather forecast service. Regarding the treatment's suggestions, we are using a simulator to generate enough data for machine learning, but the main origin of data will be the Braila's dWTP replica that we have built in AIMEN (whose main purpose is simulate critical scenarios). We are evaluating with Braila the options to install 3 sensors at the inlet of Braila's dWTP so we can show how our solution works with the real dWTP.
What is the expected size of the data?	Raw data that comes from SCADA system has a size of around 10 KB for water demand; 1 KB for noise and 2 KB for pressure. The raw weather data collected from Romanian National Meteorological Administration has a size of 88 KB and after it is been modelled the size has around 2 KB. For sensorial data, after the modelling process, the data prepared to be sent to	1Mb	For Carouge DCA, less than 5G data in total. For the IoT platform, it's not yet fully measurable because not all pilot data are collected. Currently, the development platform has 500	1MB	The data sent by CUP Braila is delivered in the form of one CSV file for each sensor, every 3 hours, each with a size of approximately 10kB. There are 4 pressure sensors, 4 noise sensors, 4 flow	To be answered by technical partners.	Since we aim to apply deep learning solutions we are expecting the order of thousands of samples. For example, if we have 5 water quality parameters and 5 weather, we need a 10x2000 matrix.

	the IoT platform could vary around 1 KB		GB capacity, and the production platform has 250GB capacity.		sensors. Other files, such as the INP files for the hydraulic models, are under 200kB		
To whom might it be useful ('data utility')?	It is useful for historical data base, for AI services and for HMI.	For policy-makers and ICT developers	Other cities, water utility companies, city workers on water related use cases.	To the Naiades use cases (Alicante, Braila, Carouge)	The data sent by CUP Braila is used for the AI noise algorithm, the AI pressure algorithm, AI weather and AI demand algorithm.	Data is useful both for the water utility (AMAEM) and for the municipality. Indirectly, it is useful for the citizens of Alicante, as the goal of the 3 uses cases is the sustainability of regional water resources	We are processing all this data, so it will be useful for us.
Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of standard identification mechanism (e.g., persistent and unique identifiers such as Digital Object Identifiers)?	The modelled data by DCA, for each sensor in part, is identified by IoT platform data base and services through the unique identifier of the physical sensor. Each type of sensor has its own entity in IoT platform. The noise sensors also provide the coordinates from the current position because they can be relocated. The rest of the sensors are placed in fixed positions and the mapping between the sensors and their addresses are known by the AI services and IoT platform based on the unique identifier of the sensor entity.	All the data have their corresponding link to the source. The information is locally stored in the webpage.	Yes. All the data has its identification.	Yes. However, some of the data are intentionally assigned a duplicate ID, which forces the old data to be discarded.	The data collected and sent by CUP Braila is standardized by SIMAVI to fit the needs of the project.	I think it is not (to be confirmed by technical partners). The data produced in Alicante's use cases are not meant to be open.	The metadata included in the data models is related to the location of the fountain and the WTP and the collection time. Currently, there is no need of more metadata since nothing else is monitored in the fountain or WTP. But they could be easily introduced.
What measures are being taken for data security (data recovery etc)?	DCA integrates an authorization and authentication mechanism to send the sensor data to the IoT platform. Also, Marketplace will integrate the same authorisation and authentication mechanism to consume IoT data. The security mechanism is using the HTTPS protocol. The data sent from DCA and Marketplace to IoT will be signed by KSI signatures. The data from the Identity Management (IDM) and Data	N/A (not necessary)	FIWARE Keyrocks and Wilma for authentication. For the development platform, no data backup is provided, but the production platform will provide data backup.	Data are being stored on Naiades platform.	All data sent by CUP Braila is backed-up within the company and can be retrieved at any time.	To be answered by technical partners.	The WTP replica data is stored locally. Nothing else is done.

	Repository could be recovered just if previously a data base back up was made.						
Are some national/ sectorial/ departmental strategies for data management used?	The data management is handled by IoT Firewall and Data Management framework.	N/A (not necessary)	N/A	N/A	The data collected by CUP Braila is in accordance with the company's own internal data management strategy.	Data is provided only on a "push" scheme; no direct access to the original data sources (SCADA, smart water meter database) is granted.	No
What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?	The access to the platform where the data base is installed is private and can be accessed only by the administrator of the platform via a private key. In the platform, security certificates will be installed for HTTPS communication. Also, the KSI signatures and KSI Gateway will be installed to validate the data integrity that circulates through the platform and for the data which comes from external clients (DCA, HMI, Marketplace).	N/A (not necessary)	NAIADES Security modules provide data security and authentication.	N/A	All data collected by CUP Braila is backed-up to local, secure, servers and no sensitive data is transmitted to the project.	To be answered by technical partners.	The WTP replica data is not confidential. The data processing of the fountain is directly collected for training from the platform but it is never stored.
Is the data safely stored in certified repositories for long term preservation and curation?	Yes, the data is saved on data base servers (PostgreSQL).	N/A (not necessary)	Data repository and safety are to be provided by NAIADES security modules.	N/A	CUP Braila maintains self-hosted long-term data storage repositories.	To be answered by technical partners.	The data we produce is not meant to be preserved for long time. Once the models are working, the outcomes will be stored in NAIADES platform repository.
What naming conventions do you follow?	N/A	Schema.org naming convention.	Schema.org.	Schema.org.	The naming conventions for the data supplied by CUP Braila to the project are determined by SIMAVI.	To be answered by technical partners.	The data is named based on the data models generated in the project.
Will search keywords be provided that optimize possibilities for re-use?	Possible for DCA in terms of collecting raw data from sensors when some of them are replaced.	Schema.org naming convention.	Yes	N/A	Not within the scope in which CUP Braila is involved in the project.	To be answered by technical partners	N/A
Do you provide clear version numbers?	N/A	GitHub (https://github.	Yes	Yes	Not within the scope in which CUP Braila	To be answered by technical partners.	N/A

		com/aolite/naiadesStandardization)			is involved in the project.		
What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.	For each modelled data by DCA metadata exists.	Schema.org naming convention.	Metadata on Flowerbed watering, weather, water quality, water consumption, etc.	N/A	Not within the scope in which CUP Braila is involved in the project.	To be answered by technical partners.	N/A
Which data produced and/or used in the project will be made openly accessible as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual	There is no decision taken by the consortium regarding the sets of accessible public data. This data will be accessed by external public users through Marketplace module.	All data is openly accessible.	The weather data is open data, but other data related to the pilot use cases are accessible by authenticated users. The collected data and service data are targeted to the specific user groups (city workers and decision makers).	N/A	Not within the scope in which CUP Braila is involved in the project.	Not fully discussed; in any case, data for UCA3 belongs to the municipality, and it cannot be shared.	N/A
How will the data be made accessible (e.g., by deposition in a repository)?	The data will be accessed through HMI and Marketplace.	It is available through the webpage: https://aolite.github.io/naiades Standardization /#/.	Through NAIADES user applications and marketplace.	Through HMI user applications provided by Naiades.	Not within the scope in which CUP is involved in the project.	To be answered by technical partners.	The predictions and suggestions will be sent to NAIADES data manager to store in the historical repository.
Which data format do you use (e.g., CSV)?	The DCA uses CSV files to read the raw data from sensorial system and JSON for the communication with the IoT platform. Also, the Marketplace module uses JSON format for communication with the IoT.	JSON	JSON & JSON-LD	JSON	CUP Braila supplies CSV data files to SIMAVI for the Naiades Project and INP files directly to the Naiades Project.	CSV for FTP export. Further information should be provided by the technical partners.	Sent in JSON format.
Is documentation about the software needed to access the data included?	For the Marketplace module, the APIs will be described in the UI API documentation page.	GitHub (https://github.com/aolite/naia	All the software documentation is provided in GitLab to the partners.	N/A	The hydraulic model, which requires an open-source program named EPANET, includes a	To be answered by technical partners.	N/A

		desStandardization).			documentation which is freely available on the internet.		
Is it possible to include the relevant software (e.g., in open-source code)?	Yes, the used technologies are described in specific deliverables. (D7 4 Architecture of the NAIADES Marketplace - Mid-term; D3 12 Communication Platform WMS).	GitHub (https://github.com/aolite/naiadesStandardization).	Yes. It's modular and open-source based.	Yes	EPANET is freely available and its source code is in the Public Domain.	To be answered by technical partners.	N/A
If there are restrictions on use, how will access be provided?	The data from IoT Repository can be accessed only through HMI and Marketplace using the authentication page based on user's role and credentials. The raw data of DCA cannot be accessible from outside because the DCA Repository is private.	N/A (not necessary)	The access is provided to the authenticated authors.	N/A	There are no restrictions.	To be answered by technical partners.	N/A
Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating recombinations with different datasets from different origins)?	Some data can be accessible through Marketplace by other institutions.	It is available through the webpage: https://aolite.github.io/naiades Standardization /#/ Also, the code an everything is openly available at: GitHub (https://github.com/aolite/naiadesStandardization).	Any NGSIv2 and NGSI-LD compatible data will be interoperable.	N/A	Not within the scope in which CUP is involved in the project.	To be answered by technical partners.	Yes, NAIDES project bases its interoperability in FIWARE, so the new users should be FIWARE compliant too.
What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?	N/A	Schema.org.	NGSIv2 and NGSI-LD	Schema.org.	Not within the scope in which CUP is involved in the project.	To be answered by technical partners.	Based on FIWARE.
Will you be using standard vocabularies for all data types present in your data	DCA will use standard vocabularies for all data types sent to the IoT platform through Data Validation modules.	Schema.org.	Yes. It's following FIWARE and ETSI standard.	Yes	Not within the scope in which CUP is	To be answered by technical partners.	N/A

set, to allow inter-disciplinary interoperability?					involved in the project.		
What is the long-term strategy for preservation of data, taking into account cost for storage and value of data?	The data storage size will be analysed, and based on that, the cloud performance could be improved.	For now, it is available at GitHub without maintenance cost during NAIADES.	The data will be stored following CA, GDPR policy and city's needs after the project's period.	N/A	Not within the scope in which CUP is involved in the project.	To be answered by technical partners.	N/A