Data Management Plan

Initial Version

AUTHORS : CHRISTOS MAKROPOULOS, GEORGE KARAVOKIROS, KLIO MONOKROUSOU

DATE : 27.12.2018

Technical References

Project Acronym	NextGen
Project Title	Towards a next generation of water systems and services for the circular economy
Project Coordinator	KWR
Project Duration	48 months

Deliverable No.	D7.2
Dissemination level [1]	PU
Work Package	WP7
Task	T7.3
Lead beneficiary	KWR
Contributing beneficiary(ies)	CTM, FHNW, UCRAN, STRANE, ECSI, NTUA
Due date of deliverable	31/12/2018
Actual submission date	27/12/2018

[1] PU = Public
PP = Restricted to other programme participants (including the Commission Services)
RE = Restricted to a group specified by the consortium (including the Commission Services)
CO = Confidential, only for members of the consortium (including the Commission Services)

Document history

V	Date	Beneficiary	Author
1	13.12.2018	KWR, NTUA	First draft by Christos Makropoulos, George Karavokiros, Klio Monokrousou
1.1	17.12.2018	CTM, FHNW, UCRAN, STRANE, ECSI, NTUA	Inputs by WP leaders on specific data provisions of the WPs
2	21.12.2018	FHNW	Quality Assurance by Anders Nättorp
3	27.12.2018	KWR	Final report by Christos Makropoulos

Summary

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the NextGen project with regards to all the datasets that will be generated by the project.

The DMP should address the points presented in this document on a dataset by dataset basis and should reflect the current status of reflection within the consortium about the data that will be generated, collected, stored and processed. Furthermore, a planning in data treatment is important for addressing timely security, privacy and ethical aspects, a process increasingly important in a multi-partner research.

In general terms, the research data should be 'FAIR' that is findable, openly accessible, interoperable and re-usable through clarifying licenses. The allocation of resources for the data collection and management is explicitly described in chapter 3 of this document. With regard to the data security, developing and establishing a backup and recovery plan for all NextGen data is a crucial part of data security. With regards to the ethical aspects, a number of ethical issues have been identified for the execution of NextGen, specifically humans, non-EU countries, data protection and compliance with the GDPR and others.

The benefits of a well-designed DMP not only concern the way data are treated but also the successful outcome of the project itself.

Disclaimer

Table of Contents

Abbreviations

CE	Circular Economy
BY-NC	Creative Commons, Attribution + Non-Commercial
CC-BY	Creative Commons, Attribution
CC-SA	Creative Commons, Attribution + Share Alike
CoP	Community of Practice
DMP	Data Management Plan
KPI	Key Performance Indicators
TEB	Technology Evidence Base
WP	Work Package

1. Data summary

State the purpose of the data collection/generation

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the NextGen project with regards to all the datasets that will be generated by the project. The DMP is not a fixed document, but evolves during the lifespan of the project; in fact, it functions as a dynamic document of agreements. The DMP should address the points presented below on a dataset by dataset basis and should reflect the current status of reflection within the consortium about the data that will be generated, collected, stored and processed.

In principle, publicly funded research data are a public good, produced for the public interest that should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property. On this basis, the DMP intends to help researchers consider at an early stage, when research is being designed and planned, how data will be managed during the research process and shared afterwards with the wider research community.

The benefits of a well-designed DMP not only concern the way data are treated but also the successful outcome of the project itself. A properly planned DMP guides the researchers first to think what to do with the data and then how to collect, store and process them, etc. Furthermore, a planning in data treatment is important for addressing timely security, privacy and ethical aspects. This way the research data are kept in track in cases of possible staff or other changes. The DMP can also increase preparedness for possible data requests. In short, planned activities, such as implementation of well-designed DMP, stand a better chance of meeting their goals than unplanned ones.

The process of planning is also a process of communication, increasingly important in a multi-partner research. The characteristics of collaboration should be accordingly harmonised among project partners from different organisations or different countries. The DMP also provides an ideal opportunity to engender best practice with regards to e.g. file formats, metadata standards, storage and risk management practices, leading to greater longevity and sustainability of data and higher quality standards.

Ultimately, the DMP should engage researchers in conversations with those providing the services. In this context, the DMP becomes a document in accordance with relevant standards and community best practice. Data should be shared, edited, and monitored among those contributing to the project. Releasing research data should follow legal, ethical and commercial terms and conditions. To serve the multiple purposes just described, the DMP should be designed for easy digital exchange across a variety of applications. The best way to approach this in today's complex world of information technology is through a metadata standard describing a data model of elements constituting the DMP.

Explain the relation to the objectives of the project

NextGen is aiming to demonstrate technological, business and governance solutions for water in the circular economy and bring new market dynamics throughout the water cycle at 10 high-profile, large-scale demonstration cases across Europe. In order to accomplish that, it is necessary to develop practical approaches, tools and partnerships as well as policy and legal framework conditions, which will accelerate acceptance of these technologies and will broaden the market reach and uptake.

The project, thus, builds on **ten Large Scale Demonstration Cases** where next generation circular water systems and services are either already deployed at full scale (six cases) or are to be deployed in the near future (four cases), demonstrating the replication potential of the approaches in practice. These real-world demonstrations allow us to create and disseminate a comprehensive **Technology Evidence Base (TEB)** (task 1.1) of the proposed solutions' effectiveness, performance, enabling conditions and opportunities for business development (**WP1**) fed by multi-parametric technology's assessment, process redesign, and long-term, uncertainty-aware, stress testing results both at the individual and whole system levels (**WP2**). Crucially, *NextGen* empowers citizens and stakeholders to learn about and be involved in the design of circular solutions from the start supporting their engagement through novel means, to both allow them to understand what 'closing the cycle' could mean in a 'risk-free' learning environment (through Serious Game) and to visualise hidden aspects of the cycle (e.g. energy flows) as part of an enhanced demonstration experience (through Augmented Reality applications) (**WP3**). The project addresses the socio-political, regulatory and governance (pre)conditions for circular water systems and services and proposes an actionable EU Roadmap for Water in the CE (**WP4**). All this work gears towards the development and demonstration of new CE business models and market opportunities, also supporting the emerging CE business ecosystem through a water in the CE MarketPlace, firmly embedded into, and extending the outreach of WssTP's real face-to-face water marketplace (**WP5**). Last but not least, *NextGen* connects, creates synergies with, and supports learning (**WP6**) and also networking with three Associated Partners as a specific case from overseas markets (India, South Korea and China) in partnership with WssTP and experts in science communication.

A dedicated *NextGen Toolkit* (hereafter termed the *NextGen Interactive Interface*), will also be developed (task 2.4) that will support decision-making through presentation of the appropriate tools for individual solution assessment and system-wide evaluation stress testing. It will also provide user-friendly access to the TEB with information of e.g. outputs of the models, databases, KPIs, relationships based on user experience, etc.

Regarding the *NextGen Marketplace* (task 5.2), several services will be developed such as linking to the NextGen Evidence Base (WP1) and allowing users to explore the showcases, finding collaborators and clients by building on a 'match-making' functionality, providing personalised information to interested parties, funding innovation and supporting spinoffs.

To improve the visibility and commercialization of technologies and circular economy solutions and systems around resource use in the water sector, a general strategic approach is developed that includes target group assessment and definition of specific sets of knowledge products for case sites and technology application (WP2).

NextGen Marketplace intends to share experiences and outcomes with stakeholder groups through its online platform that will be linked to existing networks, including WssTP water market and Project Ô.

Specify the types and formats of data generated/collected

Within WP1, data regarding the performance of the technical solutions proposed in the project will be collected as in-situ data by the site leaders. Additionally, a short description of the demo sites and some KPIs in terms of water, energy and materials will be requested by the CTGs for setting the baseline and quantifying the impact of the proposed system. A general description of each site will be collected through a fact-sheet template, where a description of the current and NextGen systems, in terms of technologies used and volumes treated, as well as block diagrams will be included. Data for KPIs will be requested from the site leaders in an excel form, which will be updated during the project execution.

WP2 will assess economic and environmental performance of individual CE water technologies and associated risks, optimise the systems and make assessment and modelling results and tools available in a sustainable manner.

In order to perform this, the following data will be collected:

1. technology specific data (e.g. system configurations, mass balances, energy and chemical demand, performance information, costs, etc.) for baseline (i.e. status quo) and NEXTGEN configuration
2. infrastructure data (storage capacity, pipes sizes, etc.) for modelling the components of the physical and engineered system
3. historical data (water and material flows, energy, etc.). Historical data will be retrieved and converted into an interoperable and standard data format (such as OGC standards including WaterML 2.0 and InfraGML) and stored for further use in the water tools Hydroptim and UWOT.
4. definition of the management boundary conditions: criteria and objectives, current and future scenarios. Management boundary conditions will be defined with water authorities/utilities and stakeholders within the CoP activities at the CS sites
5. data regarding the stakeholders need and interest for the NextGen toolbox will be collected as part of the CoP meeting. The main interest on the tool, the use cases for using it as well as the expected view of the results will be reported in an anonymized way. Once the tool is developed, it will allow the creation of user profiles giving the possibility to use the tool in a personalized way. Finally, the users will be able to provide their feedback on the returned results of the tool, aiming to collect this data and further improve the returned results.

The following data will be generated:

- Assessment of environmental impacts in various categories (e.g. global warming potential, primary energy demand, human toxicity, …) of the applied technologies using defined LCA indicators.
- Assessment of cost in standard categories (raw material cost, personnel, capex, etc.)
- Risk assessment will generate site specific uncertainty estimations to assess the probability of achieving the predefined objectives, e.g. reduction of microbial counts, compliance with of environmental legislation using recycling products. The data generated will be in the form of probability distributions as a function of system parameters (e.g. treatment time, yield)
- Model configurations combining inputs (standard or shocks including CC), systems (e.g. upstream changes, failure of components) with corresponding performance (e.g. circularity indicators of the water utilities, ecoefficiency, reliability, contamination risk, materials produced, profits, ecosystem services etc.).

Within WP3, data regarding stakeholders' views of the demonstrated circular water solutions will be collected as minutes of CoP meetings. The perspectives of the stakeholders will be reported in these minutes, e.g. as anecdotal stories. In these minutes, all personally identifiable data will be anonymised. For the purpose of cross-fertilisation between the demo cases, the CoP reports will be stored and made available to the project partners in the NextGen SharePoint.

WP4 is expected to collect and process information regarding the socio-political and governance context for water in the circular economy. Relevant information, in the form of legal documents, reports, policy papers etc., will be collected and, where appropriate, will be included in the NextGen Interactive Interface as part of the project's knowledge environment.

One main component of WP5 is the Marketplace, aiming at providing information on showcases collected in WP1, storing information of new initiatives and spin-offs in the field of innovation funding and, most important, collecting information referring to end users, technology providers and products. All information will be stored in online databases. Selected information may be exported in other forms such as pdf (factsheets) or CSV (tables).

Finally, WP6 will collect information, as above, for the project's associate partners (India, South Korea and China) and their relevant cases, technologies and contexts.

Specify if existing data is being re-used (if any)

Within WP2, and more specifically as part of the NextGen Toolkit, data about the technologies of NextGen will be used and returned to the users. Furthermore, the users will potentially have access to the data of the Technology Evidence Base that will be created as part of WP1.

The Marketplace (WP5) will build on the technology developed in the SUBSOL project. Special attention will be given in the collaboration between NextGen and Project O, to ensure synergies between the proposed Marketplace development of the two projects are exploited with a view to produce, if possible, an integrated Marketplace.

Specify the origin of the data

Data collected within WP1 will be obtained from in-situ monitoring of the performance of the technical solutions tested within NextGen.

Data collected within WP2 will be obtained from in-situ monitoring of the performance of the technical solutions tested within NextGen and from interviews with the contact persons of the sites and from the CoP of the sites as the users of the NextGen Toolkit.

Data collected within WP3 will be obtained from stakeholders participating in CoP meetings, and the general public voluntarily participating in outreach events, serious games and augmented reality activities at the demo cases. We will ensure fully informed understanding of the implications of participation through information sheets.

In WP5 data will be collected from project partners in the first phase of the project. Additional information may be collected or accessed online by connecting the Marketplace with other established European platforms. In the final phase the Marketplace will be opened to registered users which may add further content.

State the expected size of the data (if known)

In WP2 there will be data of about 10-15 participants at the CoP for the NextGen Toolkit. Furthermore, few hundreds of users are expected to use the Toolkit and provide data.

In WP3, the size of data collected is difficult to estimate but anyway substantial, as it relates to the input from approximately 4 to 6 stakeholder organisations from the ten CoPs, and thousands of citizens participating in the living labs events and three serious games / augmented reality activities.

In WP5 it is expected that the Marketplace database will store active users, technology providers and technologies in the order of hundreds.

Outline the data utility: to whom will it be useful

Data collected within WP1 will be useful for the work of other WPs of the project, such as WP2 and WP5. WP1 comparisons of the different solutions tested will be used for engineering, water utilities and other specialized audience for selecting the best circular technologies to be applied in future projects.

The data collected and generated in WP2 will be used to assess economic and environmental performance of individual CE water technologies and associated risks, optimise the systems and make assessment and modelling results and tools available in a sustainable manner.

Data collected within WP3 will be useful for the partners in the demo cases. The outcomes of the stakeholder collaboration in the CoPs and the public engagement in the Living Labs will benefit the devise, implementation and acceptability of the CE water technologies. Data collected within WP3 will be useful for WP4 of the project as well. Generic lessons from WP3 will be included in the EU roadmap for water in the CE.

2. FAIR Data

In general terms, the research data should be 'FAIR' that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard or implementation-solution.

2.1 Making data findable, including provisions for metadata

Outline the discoverability of data (metadata provision)

All NextGen subsystems, including the Technology Evidence Base, the Marketplace and the Toolkit developed within the framework of the project, will be integrated into a common online system. They all give (open or limited) access to certain data categories. Thus, one way to increase the discoverability of the data is to develop HTML pages rich in <meta> tags which provide metadata about the HTML document and are parsable by other machines. Additionally, all subsystems, and mostly the integrated system of the three elements, will be linked to the project website. Efforts will be taken to establish also cross-links to other relevant websites.

Time series collected at various project sites will be associated to the related case study areas stored in the TEB and thus will be discoverable also through this system.

NextGen will encourage project partners to store datasets produced in open research repositories and obtain Digital Object Identifiers (DOIs).

Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?

Objects stored in the NextGen TEB will be associated with a unique and persistent ID, at the appropriate resolution (i.e. scale or aggregation level). The detail page of these objects can be accessed through a specific URL, where <data category> is the data category to which this object belongs.

At the end of the project a snapshot of the whole database will be exported and stored in the research data repository Zenodo obtaining a DOI. Other categories of data produced by this project such as publications and presentations will also be uploaded to Zenodo or another research repository providing DOIs.

Outline naming conventions used

The proposed naming convention for datasets in the NextGen project will consist of three parts:

- A prefix indicating the data category (e.g. metadata, backup data, template)
- A short name for the dataset

- An integer indicating the version of the dataset

The above information will also be used as a unique code for the dataset. Additional metadata such as the data provider and the creation date will be associated with the code and uploaded to the repository.

Outline the approach towards search keyword

The NextGen Marketplace will provide a variety of ways for data retrieval, such as full text search with approximate string matching and innovative advanced search capabilities based on the attributes and relations of various data categories, including keywords and tags. Additionally, georeferenced data (e.g. case studies) will be projected on a map enabling a geographical approach for the identification of these data and the navigation to detail pages. Other platforms which this project will use for certain data categories (e.g. Zenodo, GitHub) provide similar search capabilities.

Outline the approach for clear versioning

Various information systems will have to be developed during the lifetime of the NextGen project. Many of them will require the collaborative development of software applications, websites and other computer programs either from scratch or at least significant parts of them. The management and version control of the source code may become a significant problem, especially in large software development projects involving several developers.

The NextGen IT team will systematically use **Git** as the source code management system for software development. Git will allow the collaborate development of the software providing effective distributed version control. It is easy to use and yet powerful and efficient to handle even large projects.

A copy of each Git repository will be uploaded ("pushed") to a publicly available Git server such as GitHub or Bitbucket. They are both well-established online servers which support distributed source code development, management, and revision control. They enable world-wide collaboration between developers and also provide some additional facilities to work on documentation and to track issues. GitHub provides paid and free service plans. Free service plans can have any number of public, open-access repositories with unlimited collaborators. Private, non-public repositories require a paid service plan while Bitbucket allows private repositories to be shared with up to five collaborators.

In general, all datasets produced by NextGen will receive a unique code, part of which will specify the version of the dataset.

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

Presently there are numerous standardisation activities including risk assessment and protection of a high level of services, although sometimes optional, is considered integrated in the entire operational process.

Other related Standards include:

- ISO 9001:2015 provides an integrated approach to quality management - putting quality at the heart of business, touching on business resilience.
- ISO/IEC 27001 is an internationally recognised best practice framework for an information security management system.
- ISO/TR 37150:2014 Smart community infrastructures - Review of existing activities relevant to metrics addresses community infrastructures such as energy, water, transportation, waste and information and communications technology (ICT), focuses on the technical aspects of existing activities which have been published, implemented or discussed. Economic, political or societal aspects are not analysed.
- ISO 19115, Geographic Information – Metadata
- ISO 19110 Geographic information – Methodology for feature cataloguing
- ISO 19139, Geographic Information – Metadata -Implementation Specification
- OGC Catalog Service

Efforts will be made that all textual information collected in the NextGen project and especially all new information produced and stored in the NextGen databases will be UTF-8 encoded.

2.2 Making data openly accessible

Specify which data will be made openly available. If some data is kept closed provide rationale for doing so

It is expected that most of the data produced or collected in the framework of the NextGen project will be openly accessible. However, for some reasons, related with licensing issues, 3rd party data policy, and the use of personal or sensitive data, accessing may be restricted. The Marketplace is such an example, where efforts are taken to provide more targeted and personalised information to the users, utilizing advanced methodologies such as machine learning. As part of this process, information of the individual user's preferences will be collected and stored, to which only the Marketplace application will have access.

Specify how the data will be made available

Several websites will provide access to data. They can be accessed through the [project website](#) (access to project data) or the NextGen online Marketplace. Specifically, the project's online

environment (the MarketPlace linked to the Interactive Interface which will include the TEB), will form part of WssTPs infrastructure and as such will be developed and maintained beyond the project providing access to collected as well as to new material.

Additionally, at the end of the project a copy/snapshot of all relevant data as well as other data categories (publications, presentations, etc.) will be made available to open data repositories such as Zenodo.

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

The data will be made available online through the WWW service. Thus, in order to discover, visualize or download the data only a common browser will be required from the user. Some online applications may require the user to register first before accessing the data. The format of downloadable files will be (solely or additionally) in an open standard such as CSV and PDF. Tools to access these files will be freely available in the internet. User guides will be elaborated and provided for online environments, as well as for any non-intuitive tools produced within the framework of the project.

Data stored in the NextGen Interactive Interface environment will additionally be available through a REST API for direct reuse by other systems.

Specify where the data and associated metadata, documentation and code are deposited

Several websites will provide access to data such as the following:

- the project website (Access to project data)
- the NextGen Marketplace (Products & Services, Stakeholders etc.), which will also provide access to the NextGen Technology Evidence Base.
- the Data Collection and Monitoring of the demo cases (Time series)

Additionally, at the end of the project a copy/snapshot of all relevant data as well as other data categories (publications, presentations, etc.) will be made available to open data repositories such as Zenodo.

Specify how access will be provided in case there are any restrictions

In order to gain access to the online Marketplace, a user will have to register to the Marketplace first and provide data regarding the user's background and fields of interest. The procedure is then simple and straightforward, involving the verification of the user data via email. Also, structured questions regarding specific technologies can be formed and a guided navigation with specific technological solutions will be provided.

2.3 Making data interoperable

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

Datasets produced by the NextGen project will be interoperable allowing data exchange and reuse. All systems will be user friendly, well documented and unless otherwise specified openly accessible. The NextGen project will follow established European metadata vocabularies, standards and methodologies.

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

A vocabulary used by this project will be adopted and adjusted, based on which a Taxonomy of Circular Water Solutions will be elaborated and made available to everyone through the research data repository Zenodo.

2.4 Increase data re-use (through clarifying licenses)

Specify how the data will be licenced to permit the widest reuse possible

Project partners intend to provide most of the data openly accessible, either immediately or after an embargo period. In these cases, the preferred license types are CC-BY (Creative Commons, Attribution) and CC-SA (Creative Commons, Attribution + Share Alike) and BY-NC (Creative Commons, Attribution + Non-Commercial).

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

Some data are already available online through the project websites. Other datasets will be uploaded to Zenodo as soon as they are complete and will be made immediately available through this service. Finally, for some data an embargo period will be necessary in order for them to be published first. It is estimated that for these data the embargo period will not last longer than until the end of the project.

It is not expected that NextGen will seek patents.

Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

Almost all data produced and used in this project are available by third parties or will be made available after the end of the project. A small part of them will still remain closed either because of licensing issues or because they refer to personal information (e.g. user profiles).

Describe data quality assurance processes

Data validation process will be finalized by the end of the project.

Some aspects that will be considered are the following. All parameters measured by the monitoring stations must be checked against extreme values in order to identify possible defect of instruments or another malfunction of the system. The responsible user will be notified in case of an issue.

Upon data entry into the Evidence Base or the Marketplace several validation routines must apply informing the user in case the record is incomplete or erroneous. Such validation checks include missing mandatory values and data type check.

Finally, several constraints embedded in the database should ensure the integrity of the data.

Specify the length of time for which the data will remain re-usable

All open data will remain re-usable for at least five years after the end of the project and maintained by the project partners. As suggested earlier, the project's online environment will form part of WssTPs infrastructure and as such will be developed and maintained well beyond the project.

3. Allocation of resources

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

Most of the project partners apply software tools, technologies and protocols which comply to the FAIR principles and/or facilitate FAIR use of data. The use of WWW for various applications of the project documents this approach. Additional effort will be required at least for the following tasks:

- Implementing various searching techniques (est. costs: 3 MM)
- Data conversion (est. costs: 2 MM)
- Establishing interoperability with other projects, WssTP water market and Project Ô in particular (est. costs: 6 MM)
- Managing data in open research data repositories (est. costs: 1 MM)
- Developing and implementing a data management plan (est. costs: 1 MM)
- Implementing special interfaces (API) for accessing data through the Internet (est. costs: 2 MM)

The costs will be covered by the project partners responsible for the respective tasks.

Clearly identify responsibilities for data management in your project

KWR is the responsible partner for the overall coordination of the data management.

CTM is responsible for developing and coordinating the collection of relevant data and populating the Technology Evidence Base.

ICCS is responsible for developing the NextGen Toolkit.

NTUA is responsible for developing the NextGen online Marketplace.

Additional data managers at the various project partner demo cases (WP1) or Communities of Practice (WP3) take care of the other data produced locally.

Describe costs and potential value of long-term preservation

A snapshot of all relevant project data will be stored in the research data repository Zenodo upon termination of the project. This dataset will receive a DOI and will be preserved for the foreseeable future. However, the dataset will continue to grow even after the end of the project as real time measurements will keep coming from the various project sites. Project partners responsible for the stations and other local stakeholders involved in their management and the exploitation of the data have an interest to keep these stations in operation as long as possible.

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

Developing and establishing a backup and recovery plan for all NextGen data is a crucial part of data security. While a general-purpose repository like Zenodo will store and publish the significant releases of the data produced in NextGen, a backup plan takes care of the daily backups of all data. A variety of backup software exists for all operating systems, enabling manual and automated backup of the data. State of the art backup systems offer efficient storage of data (full and incremental backup, differential, backwards deltas) which keep track with older versions and even allow the reconstruction of the data at any single point in time.

From the perspective of the data security, we distinguish between the following data categories:

- Data having no overall importance for the project. These data are often temporary and/or serve as an intermediate step for the production of other data. They are usually stored in various files. For this category of data, it is in the responsibility of the local system administrators to establish reliable backup plans, if necessary.

- Data stored in the project's website, in the NextGen Evidence Base the Toolbox or the Marketplace will be backed up centrally in a remote backup server. The backup server will be located physically at a different site than the other servers and will be connected with them through the Internet. Thus, it will not be affected by any malfunction or damage which may occur at the site of each server. A filesystem snapshot utility based on rsync (e.g. rsnapshot) will be used to make periodic snapshots of local servers and transfer the data securely via ssh protocol to the remote backup server.

- Data collected by the monitoring systems at the various case study sites. These data can grow into large volumes depending on the size of the monitoring network, the number of the observed parameters and the recording time step. In case of high frequent observations, more frequent backups or establishing a replication system will be required. Depending on each case, monitoring data may be and stored in local and/or remote backup servers.

- The NextGen Git repository stored in a remote server will serve as a means to backup software code produced during the project. An established service provider will be selected (e.g. GitHub or Bitbucket) guarantying long term preservation of the data. As the code is expected to be open source, its management will be taken care by the community.

At the end of the project, a copy/snapshot of all significant data will be stored in the project's repository at Zenodo. The infrastructure of Zenodo has been developed and is supported by CERN which guarantees long term data safety and availability.

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

A number of ethics issues have been identified for the execution of NextGen, specifically humans, non-EU countries, data protection and compliance with the GDPR and others. A response to each of these issues is given below.

Humans

Ethics issues in NextGen are a subject of this deliverable, other deliverables (D8.1 & D8.2) as well as the relevant framework and will continue to be developed and improved throughout the project.

As humans are involved in NextGen, ethical issues related to their involvement, notably regarding recruitment and informed consent procedures need to be sufficiently described.

In D8.1, it is explicitly described how personal data as part of workshops, focus groups, interviews, CoPs, living labs and other dissemination activities in NextGen should be collected, used and managed. These reports also propose measures to protect particular individuals and stakeholders, anonymise their contribution and minimise the risk of their stigmatisation from improper use of their data. As a rule, it is recommended that to the extent possible, only anonymized information is stored as part of the collected data.

In D8.2, the foreseen ethics considerations are explicitly described, related to the collection and processing of the data required for research and development of the NextGen project. Issues such as data collection, informed consent, confidentiality, the code of conduct and the balance between data protection and freedom of rights have been identified. Information is also provided on procedures to be implemented for data collection, storage, protection, retention and destruction in compliance to EU Legislation.

In NextGen, demo cases are being performed in close cooperation with local end-users, knowledge providers, technology providers, and water management agencies, thereby building local CoP groups. In NextGen, cooperation through associate partnerships has been established.

Data will be collected from participants in workshops and other opinion elicitation activities. Personal data may also be obtained from already existing data files owned by the different companies, industries and public administrations involved to the project, provided that said entities already poses the necessary consents to reuse personal data for research purposes. Each of the data sources used will be already complying with the data protection legislation applicable to each of the countries of origin, and in particular all data files should already be registered at the relevant data protection agency. All original data files have to be particularly

compliant with the right to access, modify, cancel and object to further treatment of personal data.

Non-EU countries

The objective of NextGen is to establish a market breakthrough of circular water technologies as robust, effective, sustainable, and cost-efficient answers to the water, energy and material challenges in urban areas worldwide. It practically explores replication potential though adoption of systems and services by demo cases within the project and opens new global markets in regions such as India, China and South Korea.

These regions were selected based on the severe water resources problems encountered and the long-term working relationships that already exist between local stakeholders and NextGen partners, particularly recognizing specific framework conditions, such as socio-economic, cultural and environmental characteristics. Addressing the social and governance challenges to uptake of circular solutions for water systems and services (WP4) and promoting dissemination and capacity building activities (WP6), encourages the sharing of knowledge with stakeholders. Addressing CE issues in water scarce urban areas, will enhance the availability of sufficient and high-quality water for ecology, agriculture, drinking water and other water-related industries, thus increasing the sustainable development of urban environments.

The consortium confirms that the ethical standards and guidelines of Horizon2020 will be rigorously applied, regardless of the country in which the research is carried out.

Data protection and compliance with the GDPR

The NextGen consortium will inform the EASME regarding each activity what data will be collected stored and processed as a result of these interactions. In those cases where personal data is involved, detailed information is planned to be provided:

1. on what personal data is collected, stored and processed
2. on the recruitment process, inclusion/exclusion criteria for participation
3. on privacy/confidentiality and the procedures that are implemented for data collection, storage, access, sharing policies, protection, retention and destruction during and after the project
4. on how informed consent is pursued
5. if application/s need to be filed with a local/institutional ethics review bodies (if personal data is being collected) and if yes, which bodies / where / when.

Data will be collected in accordance with the General Data Protection Regulation (GDPR), a Regulation describing precisely the rights of data subjects and the obligations of data controllers and processors. The GDPR affects mainly the Marketplace as this is expected to be the main component where information related to personal data is collected. Persons

registering for getting access to the Marketplace are providing their consent through the online form using clear and plain language. Users will be able to withdraw their consent at any time, get informed of the data the system has collected and stores about them and apply to export them or permanently erase them from the system. Sensitive data collected by the Marketplace will be encrypted before storing them in the database.

6. Other

Refer to other national / funder / sectorial / departmental procedures for data management that you are using (if any)

Up to now there is no other procedure for data management in the framework of the NextGen project, apart from the ones that have been already described in the present document.

7. Conclusions

Most of the data produced or collected in the framework of the NextGen project are generally expected to be openly accessible. Exceptions may apply for reasons related with licensing issues, third party data policy, and the use of personal or sensitive data, thus accessing this information may be restricted.

In particular, all NextGen subsystems, including the Technology Evidence Base, the Toolkit and the Marketplace, will be integrated into a common online system, which will be giving open or limited access to certain data categories.

Additionally, datasets produced by the NextGen project will be interoperable allowing data exchange and reuse. All systems will be user friendly, well documented and unless otherwise specified openly accessible. The NextGen project will follow established European metadata vocabularies, standards and methodologies.

All open data will remain re-usable for at least five years after the end of the project and maintained by the project partners.

A backup and recovery plan will be developed and established for all NextGen data as it comprises a crucial part of data security. A number of ethics issues have been identified for the execution of NextGen, specifically humans, non-EU countries, data protection and compliance with the GDPR and others.

In conclusion, the benefits of a well-designed DMP not only concern the way data are treated but also the successful outcome of the project itself. A properly planned DMP guides the researchers first to think what to do with the data and then how to collect, store and process them. The process of planning is practically a process of communication, increasingly important in a multi-partner research. The characteristics of collaboration should be accordingly harmonised among project partners from different organisations or different countries.