Draft Data Management Plan

Deliverable D1.2

Final version (public)

WP1 Project Management

Task 1.3 Information and Data Management

Lead beneficiary: VITO

Authors: Paul Campling (VITO), Sofie Van Ermen (VITO), Anshuman (TERI)

Approved by WP Managers	Paul Campling and Anshuman
Date of approval	31/07/2020
Approved by Project Coordinator	Paul Campling
Date of approval	31/07/2020
Due date of deliverable	31/07/2020
Actual submission date	31/07/2020

Project number (EU)	821051 (H2020-SC5-2018)
Project number (India)	BT/IN/EU-WR/39/AJ/2018
Project coordinator	Paul Campling, VITO
Scientific coordinator	Anshuman, TERI
Project website	https://www.pavitra-ganga.eu/en

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821051.
This project has been co-funded by Department of Biotechnology (DBT), Government of India.

Dissemination level		
PU	Public	X
CO	Confidential, restricted under conditions set out in Model Grant Agreement	
CI	Classified information as referred to in Commission Decision 2001/844/EC)	
Deliverable type		
R	Document, report	
DEM	Demonstrator, pilot, prototype	
DEC	Websites, patent fillings, videos, etc.	
ORDP	Open Research Data Pilot	X
ETHICS	Ethics	

History			
Version	Date	Reason	Revised by
0.1	1/07/2020	First draft	Paul Campling (VITO)
0.2	31/07/2020	Final Version	Paul Campling (VITO)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821051.
This project has been co-funded by Department of Biotechnology (DBT), Government of India.

SUMMARY

PAVITRA GANGA is participating in the Open Research Data Pilot of the European Commission and this is the Draft Data Management Plan, which will be finalised at the end of the project. The Data Management Plan (DMP) explains how data is to be handled during and after the project. The Pavitra Ganga DMP is based on answering the questions included in the H2020 Online Manual. It is expected that the document is periodically updated to reflect the data management decisions taken during the project.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821051.
This project has been co-funded by Department of Biotechnology (DBT), Government of India.

TABLE OF CONTENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821051.
This project has been co-funded by Department of Biotechnology (DBT), Government of India.

LIST OF FIGURES

No figure entries found.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821051.
This project has been co-funded by Department of Biotechnology (DBT), Government of India.

LIST OF TABLES

No table entries found.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821051.
This project has been co-funded by Department of Biotechnology (DBT), Government of India.

CHAPTER 1 DATA SUMMARY

Key questions and answers that address the data collected and generated by Pavitra Ganga – purpose and relevance, origin and types, data file formats and naming conventions:

- *What is the purpose of the data collection/generation and its relation to the objectives of the project?*

Pavitra Ganga aims to fulfill Sustainable Development Goal 6 by unlocking the environmental and economic potential of municipal wastewater treatment, re-use and resource recovery (RRR) for urban and peri-urban areas in India. This objective requires that the collection and generation of data provides evidence-based information to justify the choice of technologies, monitoring and modelling approaches and policy instruments to meet this potential.

- *What types and formats of data will the project generate/collect?*

Pavitra Ganga will collect and generate four types of data:

Data Type 1: Interviews, surveys and co-creation workshops to understand and describe the municipal wastewater treatment and re-use and resource recovery (RRR) approaches in India (WP2)). Information is used for problem identification and structuring, goal setting, responsibility sharing, opinions, policy assessments, and wastewater safety.

Data Type 2: Biophysical data to carryout the benchmark assessment of the water resources (water quantity, quality & infrastructure) in the two case areas and establish the hydrological assessment baseline (meteorology, land use and vegetation, soil, topography, hydrology) and socio-economic data (WP4). GIS layers to identify the resource units, time series meteorological & hydrological data, socio-economic inventory data on water demand, water supply and waste management.

Data Type 3: Water quality monitoring data (including surface, ground and drinking water) collected from stakeholders and generated by the project (WP4). Time series data of physical, chemical and biological indicators from georeferenced monitoring stations.

Data Type 4: Laboratory and piloting data of the different wastewater treatment and reuse and resource recovery (RRR) technologies generated by the project (WP3 and WP5). Water quality data and resource recovery information generated during the project at the laboratories in Europe and India and at the two piloting sites in India to verify performance.

- *Will you re-use any existing data and how?*

In the project we will re-use many existing data sets (Type 2 and 3) including; bio-physical data, water quality data, and socio-economic data.

- *What is the origin of the data?*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821051.
This project has been co-funded by Department of Biotechnology (DBT), Government of India.

The origin of the data is from different Government of India or State authorities or have been collected by IIT Kanpur and IIT Delhi in previous water quality monitoring campaigns.

- *What is the expected size of the data?*

The size of the data is expected to be several TB of data, depending on the finalisation of the data inventory, the data monitoring and the lab-scale & pilot-scale technology assessments.

- *To whom might it be useful ('data utility')?*

The data collected and generated will provide evidence-based information to stakeholders involved with wastewater management in India.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821051.
This project has been co-funded by Department of Biotechnology (DBT), Government of India.

CHAPTER 2 FINDABLE, ACCESSIBLE, INTEROPERABLE, REUSABLE (FAIR) DATA

2.1. MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA

Key questions and answers that address how data are found and the provisions made for metadata:

- *Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?*

Data resulting from the project will be made available and discoverable with metadata and a DOI, using the Zenodo platform (zenodo.org) and its functionalities to do so.

- *What naming conventions do you follow?*

The naming conventions for new data use the source and the date that the data is collected.

- *Will search keywords be provided that optimize possibilities for re-use?*

Yes, and to achieve that we will make use of Zenodo's support in harvesting of all content via the OAI-PMH protocol.

- *Do you provide clear version numbers?*

We will employ Zenodo DOI versioning.

- *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.*

We aim to follow the DataCite standards (www.datacite.org), with optionally additional DataCite recommended terms and Zenodo's enrichments.

2.2. MAKING DATA OPENLY ACCESSIBLE

Key questions and answers that address how data are made openly accessible:

- *Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.*

The case study managers at of the two pilot innovation sites in Kanpur and New Delhi specify which data is openly accessible and which data is closed. The data sharing between the technology providers and the respective case study managers must agree on a data sharing protocol. The case study manager at Kanpur has insisted that there is a data sharing protocol signed between the technology providers and IIT Kanpur. This has been done. The general rule of thumb is that publicly available data

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821051.
This project has been co-funded by Department of Biotechnology (DBT), Government of India.

sets used during the project are openly accessible according to the rules and regulations of the data owners. Most environmental agencies will provide ready access to publicly available data, if it is not used for commercial purposes.

- *How will the data be made accessible (e.g. by deposition in a repository)? What methods or software tools are needed to access the data?*

Data is shared within the project partners using the VITO share-point platform. The open data will be made available through deposition in the Zenodo repository.

- *What methods or software tools are needed to access the data?*

Data is shared as ASCII files, MS Excel, shapefiles or raster files.

- *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)?*

This is not relevant.

- *Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.*

In the Zenodo repository.

- *Have you explored appropriate arrangements with the identified repository? If there are restrictions on use, how will access be provided?*

We have explored the arrangements for the Zenodo repository and found these satisfactory for the purposes of our project.

- *Is there a need for a data access committee?*

No – all data matters are discussed at the monthly Project Management Team meeting.

- *Are there well described conditions for access (i.e. a machine-readable license)?*

We will follow the licensing schemes offered by the Zenodo repository.

- *How will the identity of the person accessing the data be ascertained?*

Through the Zenodo protocol that allows for an authentication and authorization procedure.

2.3. MAKING DATA INTEROPERABLE

Key questions and answers that address how data are made interoperable:

- *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 re-combinations with different datasets from different origins)?*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821051.
This project has been co-funded by Department of Biotechnology (DBT), Government of India.

All data will be in principle be interoperable, by using standard commercial data formats (e.g. MS Excel, ASCII, shapefiles).

- *What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable? Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?*

The aim is to comply with the JSON schema (json-schema.org) to represent metadata. In general, we aim to use as much as possible vocabularies that follow the FAIR principles.

- *In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?*

If possible, yes.

2.4. INCREASE DATA RE-USE (THROUGH CLARIFYING LICENCES)

Key questions and answers that address how the possibility to increase data re-use:

- *How will the data be licensed to permit the widest re-use possible?*

Data generated during the project will not be licensed – non-confidential data will be made publicly available.

- *When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.*

Data that serve as a basis for scientific publications will be embargoed until the actual publication is released.

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

All publicly available data are intended to be useable by third parties.

- *How long is it intended that the data remains re-usable? Are data quality assurance processes described?*

In principle, the data will remain re-usable indefinitely, though in practice the expected evolution of the state of the art will make most of the data obsolete after some years.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821051.
This project has been co-funded by Department of Biotechnology (DBT), Government of India.

CHAPTER 3 DATA MANAGEMENT ISSUES

Further to the FAIR principles, the Pavitra Ganga DMP addresses: the allocation of resources, data security, ethical issues and other issues.

3.1. ALLOCATION OF RESOURCES

Key questions and answers that address how resources are allocated to manage project data:

- *What are the costs for making data FAIR in your project?*

The main costs for making data FAIR will be the publishing of scientific papers in Open Access journals.

- *How will these be covered?*

Costs will be covered by the H2020 and DBT grant.

- *Who will be responsible for data management in your project?*

The project coordinator (VITO) is mainly responsible for following up the process, and the partners are each responsible for the data they contribute.

- *Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?*

We will be using the Zenodo platform, which envisions long-term storage for free.

3.2. DATA SECURITY

Key questions and answers that address how data security is managed:

- *What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?*

The Zenodo repository provides a high degree of data security, among others by creating duplicate data sets at differently located disk servers.

- *Is the data safely stored in certified repositories for long term preservation and curation?*

Yes, we will use the Zenodo repository, which provides a high degree of long-term preservation and curation.

3.3. ETHICAL ASPECTS

Key questions and answers that address how ethical aspects are dealt with:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821051.
This project has been co-funded by Department of Biotechnology (DBT), Government of India.

- *Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).*
- *Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?*

There are clear ethical issues in Pavitra Ganga as interviews and workshops are carried out with different stakeholders in the two case areas. Full ethical aspects are provided in the Pavitra Ganga D9.2 POPD – H Requirement No. 2.

3.4. OTHER ISSUES

Key questions and answers that other issues arising during the project are managed:

- *Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?*

Pavitra Ganga also follows the data management protocols expected by the co-funding agency Department of Biotechnology (DBT), Government of India.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821051.
This project has been co-funded by Department of Biotechnology (DBT), Government of India.