LOTUS

LOW-COST INNOVATIVE TECHNOLOGY FOR WATER QUALITY MONITORING
AND WATER RESOURCES MANAGEMENT FOR URBAN AND RURAL WATER SYSTEMS IN INDIA

Deliverable 8.3

LOTUS Data Management Plan

Lead: inno TSD

Date: December 2019

- Public -

Table of Contents

LOTUS is co-funded by the European Commission under the Horizon 2020 research and innovation programme under Grant Agreement N° 820881 and by the Indian Government, Ministry of Science and Technology

2

Project Deliverable

Project Number	Project Acronym	Project Title
820881	LOTUS	LOw-cost innovative Technology for water quality monitoring and water resources management for Urban and rural water Systems in India

Instrument:	Thematic Priority
Research and Innovation action	EU-India water co-operation

Title
D8.3 LOTUS Data Management Plan

Contractual Delivery Date	Actual Delivery Date
July 30, 2019 (M5)	December 16th, 2019 (M11)

Start Date of the project	Duration
February 1st, 2019 June 1st, 2019 (India)	48 months

Organisation name of lead contractor for this deliverable	Document version
Inno TSD	V1.0

Dissemination level		Deliverable Type	
Public	X	**Document, Report**	X
Confidential		Demonstrator	

Authors (organisations)
inno TSD

Reviewers (organisations)
EP

Abstract
The purpose of the LOTUS communication strategy and initial plan for dissemination is to provide a formal framework for communication about LOTUS and disseminating results.
Communication strategy will notably address the main communication objectives and desired impacts and identify the desired indicators and target audiences in both EU countries and India. The document describes how the consortium plans to utilise multiple dissemination channels and to employ general as well as specially targeted communication measures. Aside these contents, a first plan for dissemination of results will be drafted, further refining the dissemination objectives and exploiting the opportunities. It also indicates the role and responsibilities of the partners and identifies the audience and the key messages that should be spread.

Keywords
Communication plan, Communication strategy, Dissemination, outreach

Disclaimer

The LOTUS Project

LOTUS is a project funded by DG Environment under the European Union Horizon 2020 Research and Innovation Programme and by the Indian Government. It brings together EU and Indian prominent organisations with the aim to co-create, co-design and co-develop innovative robust affordable low-cost sensing solutions for enhancing India's water and sanitation challenges in both rural and urban area.

The LOTUS solution is based on an innovative sensor and includes tailor-made decision support to exploit the capabilities of the sensor as well as a specific approach to co-creation. LOTUS aims to be co-designed and co-produced in India, and have a wide, diverse and lasting impact for the water sector in India due to intense collaborations with commercial and academic partners in India.

Based on the low-cost sensor platform, solutions for the early detection of water quality problems, decision support for countermeasures and optimal management of drinking and irrigation water systems, tailored on the functionalities of the new sensor, will be developed and integrated with the existing monitoring and control systems.

This sensor will be deployed in five different use cases: in a water-network, on ground-water, in irrigation, in an algae-based waste water treatment plant and in water tankers. The packaging of the sensor, as well as the online and offline software tools will be tailored for each of the use cases. These last will enable to test the sensors and improve them iteratively.

The project is based on co-creation, co-design and co-production between the different partners. Therefore, an important stakeholder engagement process will be implemented during the project lifetime and involve relevant stakeholders, including local authorities, water users and social communities, and will consider possible gender differences in the use and need of water. Broad outreach activities will take place both in India and in Europe, therefore contributing to LOTUS impact maximisation.

The further development and exploitation (beyond the project) of the novel sensor platform will be done in cooperation with the Indian partners. This will create a level playing field for European and Indian industries and SMEs working in the water quality area.

LOTUS is co-funded by the European Commission under the Horizon 2020 research and innovation programme under Grant Agreement N° 820881 and by the Indian Government, Ministry of Science and Technology

5

1 Executive summary

This deliverable proposes the data management plan (DMP) to be supplied in the context of the H2020 pilot action on open access to research data. Usual practices in relation to data management plan have been looked at and allowed to identify data sets characteristics to be looked at and being: description of data to be generated or collected (explanation of the different types of data which will be produced, including file formats where possible), standards and metadata (presence and description of the metadata accompanying the dataset), data sharing (provided access level) and archiving and preservation (procedures ensure preservation of the dataset).

The target datasets in the context of the H2020 pilot on research data focuses on providing on-line access to scientific information that is free of charge to the end-user. In the context of R&D, 'scientific information' can refer to peer-reviewed scientific research articles (published in academic journals) and also to scientific research data (data underlying publications, curated data and/or raw data).

A template to describe the relevant project datasets is defined and includes: reference and name, data set description, metadata, quality assurance, data sharing and archiving and preservation, in-line with the H2020 recommendations.

From the analysis of available datasets, 7 have been selected as candidate for the Open Research data Pilot and cover the fields of: sense city sensors, sensor lab benchmarking, energy availability on the water network, energy harvesting system calibration, on site Energy harvester benchmarking, demonstrator dataset, demonstrator dataset and Communication technology testing.

Evolution of these dataset availability will be monitored at project reviews.

LOTUS is co-funded by the European Commission under the Horizon 2020 research and innovation programme under Grant Agreement N° 820881 and by the Indian Government, Ministry of Science and Technology

6

2 Introduction

H2020 has launched an Open Research Data pilot in 2014-2015. As of 2017, all H2020 project are allowed to take part of the pilot. This pilot aims at supporting beneficiaries in making their research data *"findable, accessible, interoperable and reusable (FAIR)"*. LOTUS has volunteered to be part of the Open Research Data pilot. This report constitutes LOTUS Data Management Plan (DMP).

"A further new element in Horizon 2020 is the use of Data Management Plans (DMPs) detailing what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. The use of a Data Management Plan is required for projects participating in the Open Research Data Pilot. Other projects are invited to submit a Data Management Plan if relevant for their planned research."

The guidelines further detail information to be provided within the DMP as explained within the University of Bristol guidelines:

- ✓ **Description of data to be generated or collected:** explanation of the different types of data which will be produced, including file formats where possible. An accurate description of the dataset to be produced from the start is difficult. It is therefore not expected to list everything which is subsequently produced, but to highlight which datasets the consortium agrees are the most significant and likely to have long-term value. It is also expected to state the origins of any collected data, for whom the data might be useful, and whether or not it will underpin a scientific publication. Highlighting whether similar data already exists and the possibilities for integration and reuse will also be helpful. For these reasons, regular updates of the DMP are expected.
- ✓ **Standards and metadata:** a specific attention has to be paid to the metadata accompanying the dataset. Such metadata includes the filenames and folder structure which need to be self-explanatory for 3rd party readers.
- ✓ **Data sharing:** The EC require clarity on the level of access that will be provided; will the data be widely and publicly open or will there be restrictions on who can access the data? If the latter is the case access procedures to be put in place need to be outlined. Also, if for any reason there is a need for an embargo period on the data you, this should be clearly stated and rationale given. Information on any software and other tools necessary for enabling re-use of datasets should also be provided.
- ✓ **Archiving and preservation:** Procedures which will ensure the preservation of your data, including backup and storage need to be specified. It is important to remember that some cloud based storage options may not be governed by EU legislation (such as the Data Protection Directive), and so would not be suitable for storing sensitive data.

These are the basic requirements for DMP and are further identified and developed, in this document in the context of LOTUS.

LOTUS is co-funded by the European Commission under the Horizon 2020 research and innovation programme under Grant Agreement N° 820881 and by the Indian Government, Ministry of Science and Technology

7

3 Scope

3.1 Description

Before detailing aspects of the DMP, it is required to clearly identify the scope of a DMP in the context of the H2020 pilot on Open Research Data. It is indeed not any data being generated or used by the project which has to be considered. As mentioned in the fact sheet to Open Access in Horizon 2020, open access can be defined as the practice of providing on-line access to scientific information that is free of charge to the end-user. In the context of R&D, 'scientific information' can refer to peer-reviewed scientific research articles (published in academic journals) and also to scientific research data (data underlying publications, curated data and/or raw data).

The notion of openness is central for the selection of research data to be included within the LOTUS participation to the pilot on open research data. Openness is defined by Open Knowledge as having the following characteristics:

- ✓ **Availability and access:** the data must be available as a whole as and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
- ✓ **Reuse and redistribution:** the data must be provided under terms that permit reuse and redistribution including the intermixing with other datasets. The data must be machine-readable.
- ✓ **Universal participation:** everyone must be able to use, reuse and redistribute — there should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

The fact sheet makes clear that Open access is not a requirement to publish, as researchers are free to publish or not, nor does it interfere with the decision to exploit research results commercially e.g. through patenting. Indeed, the decision on whether to publish open access documents must come after the more general decision on whether to go for a publication directly or to seek first protection using Intellectual Property Rights (IPR). This is depicted on Figure 1.
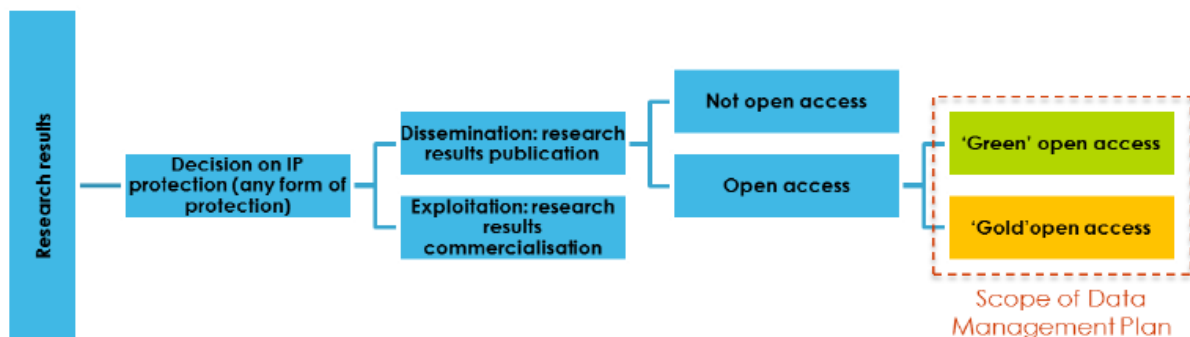

Figure 1 Open research data decision process

8

The fact sheet also defines two main and non-mutually exclusive routes towards open access to publications:

- ✓ **Self-archiving (also called 'Green' open access)** means that the published article or the final peer reviewed manuscript is archived by the researcher – or a representative - in an online repository before, after or alongside its publication. Access to the article is often – but not necessarily - delayed ('embargo period') as some scientific publishers may wish to recoup their investment by selling subscriptions and charging pay-per-download/view fees during an exclusivity period.
- ✓ **Open access publishing (also called 'Gold' open access)** means that an article is immediately provided in open access mode by the scientific publisher. The associated costs are shifted away from readers, and instead to (for example) the university or research institute to which the researcher is affiliated, or to the funding agency supporting the research.

3.2 Application to LOTUS case

The LOTUS DMP is restricted to the 2 routes described above and elaborates on the characteristics of identified data described hereunder. These are inspired from the on-line DMPTool :

3.2.1 Reference and name

An identifier for the data set has to be produced.

3.2.2 Data set description

Description of the data that will be generated or collected, its origin (distinguish between newly collected data and data being re-used from other projects), nature, scale and resolution and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.

Data types could include text, spreadsheets, images, 3D models, software, audio files, video files, reports, surveys, etc.

3.2.3 Metadata

Reference to existing suitable standards for metadata of the discipline. If these do not exist, an outline on how and what metadata will be created.

As examples, spatial data sets must be documented using either the FGDC version 2.0 or the ISO 19115 metadata standard. The Biological Data Profile standard (associated with FGDC) is very useful for creating documentation of field- and lab-based work. We recommend use of a metadata documentation tool, e.g., Metavist (http://nrs.fs.fed.us/pubs/2737).

LOTUS is co-funded by the European Commission under the Horizon 2020 research and innovation programme under Grant Agreement N° 820881 and by the Indian Government, Ministry of Science and Technology

9

3.2.4 Quality Assurance (QA)

This section describes the steps that will be used to process and ensure quality the data. Procedure need to include: data proofing and validation, including data collection, entry, transmission, and storage. Criteria related to quality assurance (e.g. documentation, calibration, validation, monitoring, versioning, etc.) should appear here.

If any, descriptive or analytical statistics to be run on the data for quality assurance should be described.

3.2.5 Data sharing

Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).

This section will include any necessary limitations to protect sensitive data as well as how to ensure security of data considering in particular the EU's Data Protection Directive.

3.2.6 Archiving and preservation

Description of the procedures that will be put in place for short and long-term preservation (data storage and backup) of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.

4 LOTUS data sets

An identification of the possible dataset of interest within the LOTUS, project has been made. All partners have been requested to identify the datasets to be produced in the context of the project. From that list, several datasets have been characterized being in-line with the scope of the LOTUS DMP as outlined in section 3. These datasets are detailed here after.

It should be noticed that no mention of scientific publication is made here. However, it is clear the as a research project, LOTUS will issue publications. These will monitored as part of the project dissemination plan.

LOTUS is co-funded by the European Commission under the Horizon 2020 research and innovation programme under Grant Agreement N° 820881 and by the Indian Government, Ministry of Science and Technology

10

4.1 Ecole Polytechnique

Reference name	Proteus' sensor caps lab characterization data
Contributing partner	EP
Description	The dataset contains the lab characterization of the sensor caps produced during Proteus project. Format of data: csv/xls
Metadata	No standardised metadata available Timestamp available within files Labelling of files and directories indicate the ID of the sensor under measurement as well as the conditions of measurement
Quality insurance	No quality insurance
Data sharing	By default, restricted access. May be partially opened to consortium partners, upon request justified with predetermined research goals
Archiving and preservation	Under investigation – current storage is within personal PC with several backups.

Reference name	Sensor chip fabrication data
Contributing partner	EP
Description	The dataset will contain the fabrication parameters of all chips fabricated during LOTUS (formulations of inks, wafer of origin of the chip, printing parameters), as well as the characterization data (various types of microscopy and spectroscopy, electronic characterization). Data generated during LOTUS Format of data: heterogeneous – text, csv, images
Metadata	Under investigation – standardisation of data management (including metadata production) is part of the short-term roadmap of EP team.

Quality insurance	Under investigation– standardization of data management (including quality insurance) is part of the short-term roadmap of EP team.
Data sharing	By default, restricted access. May be partially opened to consortium partners, upon request justified with predetermined research goals
Archiving and preservation	Under investigation – current storage is within lab-level cloud.

Reference name	**Sensor chip lab characterization data**
Contributing partner	EP
Description	The dataset contains the lab characterisation of the sensors produced within LOTUS context. Data generated during LOTUS Format of data: text, csv, xls
Metadata	Under investigation – standardisation of data management (including metadata production) is part of the short-term roadmap of EP team.
Quality insurance	Under investigation– standardisation of data management (including quality insurance) is part of the short-term roadmap of EP team.
Data sharing	By default, restricted access. May be partially opened to consortium partners, upon request justified with predetermined research goals
Archiving and preservation	Under investigation – current storage is within lab-level cloud.

4.2 AUTARCON

Reference name	Proteus' sensor caps lab characterization data
Contributing partner	AUTARCON
Description	The dataset contains the lab experiment and pilot system testing in lab. Including chlorine concentration (from proteus sensor), chlorine production rate and electrolytic cell operation characteristics. Data will be taken manually and stored in an Excel sheet. During the real case scenario, the data will be recorded online and transferred via GSM or LoRa. The online data will be stored on AUTARCON servers. Format of data: xls
Metadata	No standardized metadata available Every online dataset will receive a time stamp.
Quality insurance	No quality insurance
Data sharing	By default, restricted access. May be partially opened to consortium partners, upon request justified with predetermined research goals
Archiving and preservation	Under investigation – current storage is within personal PC with several backups and on AUTARCON servers.

4.3 DFM

Reference name	**Digital data from sensors collected over LOTUS network, post deployment.**
Contributing partner	DFM
Description	After deployment of LOTUS sensors in various use cases, data from the sensors will be collected over the network and stored in centralized servers.
Metadata	Physical location of data source (co-ordinates). Timestamp, Node IDs, Node type (type of use-case), Communication channel (Network used to send the data).
Quality insurance	No quality insurance.
Data sharing	Opened to consortium partners.
Archiving and preservation	Centralized servers and local storage (Project PCs).

4.4 IITG

Reference name	**Proteus' sensor caps lab characterization data**
Contributing partner	IITG
Description	The dataset contains the lab characterisation of the sensor caps produced during the LOTUS project. Format of data: csv/xls
Metadata	No standardised metadata available Timestamp available within files Labelling of files and directories indicate the ID of the sensor under measurement as well as the conditions of measurement

Quality insurance	No quality insurance
Data sharing	By default, restricted access. May be partially opened to consortium partners, upon request justified with predetermined research goals
Archiving and preservation	Data will be stored in Lotus Project Sharepoint folder and Archiving feature of sharepoint will be unitised. Sharepoint →link.

4.5 JISL

Reference name	Field validation of LOTUS Sensor
Contributing partner	JISL
Description	The dataset contains the lab validation as well as field validation of the sensors of LOTUS project. Format of data: csv/xls
Metadata	No standardized metadata available Individual sensor installed in Lab and Field file will be generated
Quality insurance	With reference to current sensor available in market
Data sharing	By default, restricted access. May be partially opened to consortium partners, upon request justified with predetermined research goals
Archiving and preservation	Within personal PC with several backups.

4.6JSPL

Reference name	Flow and pressure data for hydraulic modelling
Contributing partner	Guwahati Jal Board and Indian Institute of Technology, Guwahati
Description	The data set contains, historical flow and pressure logs as well as any onsite measurements taken with the installation of sensors
Metadata	Metadata includes, time stamp, author, revision date
Quality insurance	No specifici quality insurance
Data sharing	Data would be shared within the consortium members and as required in undertaking the hydraulic analysis
Archiving and preservation	Storage in personal computer and data backup drives and project common storage facility

4.7 TU Dortmun

Reference name	Models and simulation and optimization data
Contributing partner	TU Dortmund, Process Dynamics and Operations Group
Description	Models of irrigation systems, models of wastewater treatment plants, optimization models, simulation data, results from optimization runs, experimental data from installations; format according to the software in which the models are formulated / the data is generated.
Metadata	Usually none
Quality insurance	Scientific discussion
Data sharing	Upon request and approval.
Archiving and preservation	Backup systems at the department.

LOTUS is co-funded by the European Commission under the Horizon 2020 research and innovation programme under Grant Agreement N° 820881 and by the Indian Government, Ministry of Science and Technology

16

4.8 UNEXE

Reference name	EPANET modelling inputs
Contributing partner	UNEXE
Description	The input datasets will be collected from the Indian partner and converted into EPANET input format
Metadata	Standard text input files for EPANET model
Quality insurance	No quality insurance
Data sharing	The accessibility will depend on the restrictions by the original data provided. It would be shared within LOTUS consortium for the research activities within the project.
Archiving and preservation	To be agreed with LOTUS consortium. If the data is open accessible, UNEXE is providing a free data repository Open Research Exeter (ORE) for the long-term accessibility and reusability.

Reference name	EPANET modelling
Contributing partner	UNEXE
Description	The output datasets will be produced from EPANET modelling
Metadata	Standard binary output files for EPANET model
Quality insurance	No quality insurance
Data sharing	The accessibility will depend on the restrictions by the original data provideds. It would be shared within LOTUS consortium for the research activities within the project.

Archiving and preservation	To be agreed with LOTUS consortium. If the data is open accessible, UNEXE is providing a free data repository Open Research Exeter (ORE) for the long-term accessibility and reusability.

Reference name	Algorithm of anomalty detection, FIWARE applications, and 3D visualisation
Contributing partner	UNEXE
Description	The computer codes will be developed to provide various functions in LOTUS analyses
Metadata	The codes will be written in various programming language, C++, Python, Javascript, etc.
Quality insurance	No quality insurance
Data sharing	A delicated project will be set up in Github for sharing and maintaining the programme codes developed in LOTUS.
Archiving and preservation	Open Research Exeter (ORE) will be used for the long-term accessibility and reusability.

Reference name	Publications and presentations
Contributing partner	UNEXE
Description	The dissemination material of LOTUS
Metadata	The outputs will include reports, journal and conference articles, presentation, and other types of documents produced within LOTUS. Most will be produced and shared in PDF format.
Quality insurance	No quality insurance
Data sharing	The documents will be shared via Open Research Exeter (ORE) that is public accessible.
Archiving and preservation	Open Research Exeter (ORE) will be used for the long-term accessibility and reusability.

LOTUS

LOTUS is co-funded by the European Commission under the Horizon 2020 research and innovation programme under Grant Agreement N° 820881 and by the Indian Government, Ministry of Science and Technology.

18

4.9 UTH

Reference name	Groundwater data
Contributing partner	UTH
Description	This is not defined yet and defining it is part of the ongoing work in WP3. An indicative list is provided below: The dataset will contain **Administrative Data** (Citations, Entities—laboratories, responsible parties, etc.—, Projects, Campaigns, Normatives); **Hydrogeological Data** (Wells, springs, piezometers, technical information wells, measurements (head, well abstraction…), aquifer system, hydrogeological units (hydraulic properties, geometry, etc.)) and Hydro chemical data (Physical/chemical parameters along with laboratory regulatory guidelines).
Metadata	No standardised metadata available Timestamp and location available within files
Quality insurance	No quality insurance
Data sharing	Open to consortium partners and upon request justified with predetermined research goals
Archiving and preservation	Under investigation – current storage is within personal PC with several backups.

5 Conclusion

From the guidelines, DMP should be reviewed at regular intervals. This will be synchronised with project reviews planning. Updates to the DMP will appear within the dissemination report of the project.

Reviews will go further the information presented in section 4 by answering the following questions related to the produced scientific research data:

✓ **Discoverable:** are the data and associated software produced and/or used in the project discoverable (and readily located), identifiable by means of a standard identification mechanism (e.g. Digital Object Identifier)?

✓ **Accessible:** are the data and associated software produced and/or used in the project accessible and in what modalities, scope, licenses (e.g. licencing framework for research and education, embargo periods, commercial exploitation, etc.)?

✓ **Assessable** and intelligible: are the data and associated software produced and/or used in the project assessable for and intelligible to third parties in contexts such as scientific scrutiny and peer review (e.g. are the minimal datasets handled together with scientific papers for the purpose of peer review, are data is provided in a way that judgments can be made about their reliability and the competence of those who created them)?

✓ **Useable beyond the original purpose for which it was collected:** are the data and associated software produced and/or used in the project useable by third parties even long time after the collection of the data (e.g. is the data safely stored in certified repositories for long term preservation and curation; is it stored together with the minimum software, metadata and documentation to make it useful; is the data useful for the wider public needs and usable for the likely purposes of non-specialists)?

✓ **Interoperable to specific quality standards:** are the data and associated software produced and/or used in the project interoperable allowing data exchange between researchers, institutions, organisations, countries, etc. (e.g. adhering to standards for data annotation, data exchange, compliant with available software applications, and allowing re-combinations with different datasets from different origins)?

LOTUS is co-funded by the European Commission under the Horizon 2020 research and innovation programme under Grant Agreement N° 820881 and by the Indian Government, Ministry of Science and Technology

20