This project has received funding from the European Union's Horizon 2020 research and innovation programme under grand agreement No 773715

Grant Agreement No.: 773715

Project acronym: RESOLVD

Project title: Renewable penetration levered by Efficient Low Voltage Distribution grids

Research and Innovation Action

Topic: LCE-01-2016-2017

Next generation innovative technologies enabling smart grids, storage and energy system integration with increasing share of renewables: distribution network

Starting date of project: 1st of October 2017

Duration: 36 months

D8.4 – Data Management Plan

Organization name of lead contractor for this deliverable: UdG	
Due date:	M6 - 31th of March 2018
Submission Date:	28th of march 2018
Primary Authors	Joaquim Meléndez (UdG), Roberto Petite (UdG), Brigit Nonó (UdG),
Contributors	Ramon Gallart (EYPESA), Francisco Diaz (UPC), Miha Smolnikar (CS)
Version	1.1

Dissemination Level		
PU	Public	X
CO	Confidential, only for members of the consortium (including the Commission Services)	

Deliverable reviews

Revision table for this deliverable:		
Version 0.5	Reception Date	19th of march 2018
	Revision Date	20th of March 2018
	Reviewers	Heidi Tuiskula (SIN), Ramon Gallart (EYPESA)
Revision table for this deliverable:		
Version 1.1	Reception Date	20th of March 2018
	Revision Date	25th of March 2018
	Reviewers	Heidi Tuiskula (SIN), Ramon Gallart (EYPESA)

Table of contents

Executive Summary 5

1. Introduction 6
   1.1. Objectives 6
   1.2. Contributions of partners 6
   1.3. Report structure 6

2. Data Summary 7
   2.1. State the purpose of the data collection/generation 7
   2.2. Explain the relation to the objectives of the project 7
   2.3. Specify the types and formats of data generated/collected 7
   2.4. Specify if existing data is being re-used (if any) 8
   2.5. Specify the origin of the data 8
   2.6. State the expected size of the data (if known) 8
   2.7. Outline the data utility: to whom will it be useful 8

3. 2. Fair data 9
   3.1. Making data findable, including provisions for metadata 9
      3.1.1. Outline the discoverability of data (metadata provision) 9
      3.1.2. 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? 9
      3.1.3. Outline naming conventions used 9
      3.1.4. Outline the approach towards search keywords 9
      3.1.5. Outline the approach for clear versioning 9
      3.1.6. Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how 9
   3.2. 2.2 Making data openly accessible 10
      3.2.1. Specify which data will be made openly available. If some data is kept closed provide rationale for doing so 10
      3.2.2. Specify how the data will be made available 10
      3.2.3. Specify what methods or software tools are needed to access the data. Is documentation about the software needed to access the data included. It's possible to include the relevant software (e.g. In open source code) 10
      3.2.4. Specify where the data and associated metadata, documentation and code are deposited 10
      3.2.5. Specify how access will be provided in case there are any restrictions 10
   3.3. 2.3 Making data interoperable 11
      3.3.1. Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. 11
   3.4. 2.4 Increase data re-use (through clarifying licenses) 11
      3.4.1. Specify how the data will be licensed to permit the widest reuse possible 11
      3.4.2. Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed 11
      3.4.3. 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. 12
      3.4.4. Describe data quality assurance processes 12
      3.4.5. Specify the length of time for which the data will remain re-usable 12

4. Allocation of resources. 12
   4.1. Estimate the costs for making your data FAIR. Describe how you intend to cover these costs 12
   4.2. Clearly identify responsibilities for data management in your project 12
   4.3. Describe costs and potential value of long term preservation 12

5. Data security 12
5.1. Address data recovery as well as secure storage and transfer of sensitive data..... 12

6. Ethical aspects to be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. 13

7. Other issues 13

8. Further support in developing your DMP 13

9. Conclusions 13

10. Acronyms and abbreviations 13

Executive Summary

RESOLVD project aims to join the H2020 pilot on Open Research Data (ORD). The consortium agreement reflects the common position of the consortium w.r.t. data management plan (DMP), which follows the FAIR (Findable, Accessible, Interoperable and Reusable) principles.

In accordance with the "Guidelines on FAIR on Data Management in Horizon 2020" (Version 3.0, 26 July 2016), this deliverable details:

- What data the project will collect and generate;
- Whether, and how, this data will be exploited or shared and made accessible/open for verification and re-use;
- How this data will be curated and preserved.

This is the first version of the data management plan expected for M6. The document will be updated over the course of the project, in every periodic assessment or whenever significant changes arise to include new data sets and new results sets.

1. Introduction

1.1. Objectives

This data management plan aims to guarantee replicability and benchmarking of validated results of the project; in particular, those presented in scientific publications, addressing the following objectives that will be detailed in the different sections:

- Identification of the types of data that RESOLVD will generate, including typology, origin, volume, formats and files.
- Definition of how the data will be organized and managed and documented guaranteeing the good quality of the data.
- Preparation of the storage strategy during the project execution and data preservation (repository)
- Definition of the project data policies, including issues related to intellectual property.
- Costs analysis regarding data preservation and storage.

1.2. Contributions of partners

Partner	Contribution
UdG	Owner and editor of the document. UdG has also contributed in the sections referred to the institutional repository that is proposed for storing the data.
UPC	UPC has contributed to specify the data information regarding the data that will be generated in the laboratory, at UPC.
EYPESA	EYPESA has contributed to specify the information for the case of data generated in the real low voltage grid collected by a SCADA system, through the remote terminal units (RTU), aggregator concentrator links and the Data Concentrator Units (DCU) and smart meter.
CS	CS has contributed to report information related with possible data being generated by distributed sensing elements to be deployed in the grid.

1.3. Report structure

The preparation of this deliverable has consisted in answering the questions requested in the Guidelines on FAIR on Data Management in Horizon 2020. Core sections in D8.4 are organised as follows:

- Section 2: it describes in detail the summary of the data that will be generated/collected in the project, explaining the purpose of the data generation/collection, its relation with the objectives of the project, including also its origin, types and formats, etc.
- Section 3: it includes all the questions referred to make the data findable, accessible, interoperable and re-usable through the proposed repository.
- Section 4: it explains the allocation of resources and cost foreseen for data preservation.
- Section 5: it addresses data recovery as well as secure storage and transfer of sensitive data.
- Section 6: it will consider the ethical issues taking into account regarding sensitive information, if it's the case.

2. Data Summary

2.1. State the purpose of the data collection/generation

The data being used in the RESOLVD project will be oriented to improve knowledge on how power flow behaves in the low voltage in presence of distributed renewable generation and high variability on demand. The general purpose of the project RESOLVD is to act (schedule and control) on the low voltage grid in order to increase efficiency. With this aim, data will serve in the following purposes:

1. Enhance grid observability when monitoring: improve knowledge on demand/generation profiles, power flow computation, etc.

2. Modelling demand and generation for forecasting purposes: training of machine learning algorithms to forecast demand and generation in specific points of the grid.

3. Test and performance evaluation of both, technologies developed as part of the RESOLVD solution, and computation of KPIs during project validation: validation of proposed solution and quantification of improvements based on indicators.

2.2. Explain the relation to the objectives of the project

The overall objective of RESOLVD is to improve efficiency and the hosting capacity of distribution networks in a context of highly distributed renewable generation by introducing energy flexibility and control by acting on the grid. The following document describes dependencies of data being collected and generated with specific objectives derived from this primary one:

- Design, develop and test new hardware for monitoring and acting on the grid: These new devices will acquire physical measures of power and energy that will be used for monitoring and control.

- Resilient and efficient scheduling and operation of the LV grid: data from specific point of the grid will be used for modelling demand and generation and further forecasting.

- Analyse potential business models: data from different sources will be used to compute KPIs to be used in cost-efficiency analysis of different business models.

2.3. Specify the types and formats of data generated/collected

The type of data collected within the project is experimental coming from both laboratory and a real scenario provided by UPC and EYPESA respectively. In the case of data generated in the laboratory, at UPC premises, data refers to time series of active and reactive power exchanged by the power electronics device (energy router, ER and the state of batteries (level of charge, and voltage level). Data is acquired through a test platform where the devices are attached to.

In the case of data generated (see section 6 for ethical issues) in the real low voltage grid provided by EYPESA, the data will be collected by a SCADA system, through the remote terminal units (RTU) or / and aggregator concentrator links connected to smart meters. They include electrical variables such as active power, reactive, power, apparent power, voltage and currents and will be delivered as tables in *.csv files extracted from SCADA data base. The real environment will have also access to data collected by the metering infrastructure through Data Concentrator Units (DCU) and smart meter. This data set includes energy (active and reactive), voltage and/or current acquired periodically and updated daily to the control center. This data is associated to the metering point with an identifier known as Universal Code of Supplier Point (Spanish acronym: CUPS) and the format should be in *.xml or *.csv. It would be provided form the smart metering

data base. Also data from other instruments (PMUs and PQM) being used in the project can also be included in the plan but at this moment nature and availability of them is unknown.

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

Weather data is also relevant for either electricity generation (PV generation) and consumption forecasting. The *Servei Meteorològic de Catalunya* (Catalan Weather Service) has been contacted and it has already provided hourly data from January 1st 2008 to December 31st 2017 of the solar irradiation, temperature, wind speed and direction and humidity registered by the automatic weather stations placed in villages of Gurb and Orís. These two weather stations are the stations with irradiation measurements, which are closer to the pilot of the project. Nevertheless, the data provided by the Catalan Weather Service is not allowed to be open for the project.

The agency has provided the data exclusively for research purposes within the project. Access rights are directly managed by the Catalan Weather Service.

In case of re-using other existing data (open Access or other) datasets during the project execution, they will be included and described in further updates of this data management plan. At the moment of submission of this document (M6), it was not foreseen the use of other existing data.

2.5. Specify the origin of the data

Data described in subsection 2.3 are generated within the project. In the case of data collected in the laboratory, it will be gathered by data acquisition systems already installed in the laboratory. Measurements refer to electrical magnitudes. In the case of real low voltage grid provided by EyPESA, the data will be obtained from SCADA and the metering infrastructure through the Meter Data Collector (MDC) systems. As mentioned other instruments as PMUs and PQM could be deployed in the grid during the project.

Weather data is directly provided by the Catalan Weather Service.

2.6. State the expected size of the data (if known)

Data generated in the laboratory, at UPC, the expected size of the data does not exceed few MB in volume in the format of Excel files.

Data from the real low voltage grid scenario, has two main sources with different time resolution: Data coming from smart meters in the validation area generate 75 kB every day with a granularity of 60 minutes. On the other hand, data collected from the SCADA does not exceed 200 kB every day with a granularity of 3 minutes. Representative data is supposed to cover one year representing a volume of around 100MB (275kB x 365= 100375kB). PMUs and PQMs can supply registers at higher sampling frequency but at this time details are not available.

2.7. Outline the data utility: to whom will it be useful

The datasets generated in the project could be useful for those electricity actors and stakeholders (DSO, aggregators, technology and R+D providers, etc.) that have interest in the low voltage energy management and business models involving LV network operation (distributed resources generation, energy islands, aggregation, demand response, etc.). Data will be also useful for scientists to check theoretical results and test algorithms.

3. Fair data

3.1. Making data findable, including provisions for metadata

3.1.1. Outline the discoverability of data (metadata provision)

The metadata standards proposed to describe the dataset will be the Dublin Core and Datacite Schema, as they are a flexible and common used standards and are also the ones adopted by the European Open AIRE repository.

3.1.2. 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?

Data will be make open through AIRE compatible repositories. Identification of data in such repositories is given by unique and persistent HANDLE. For example, the UdG institutional repository "https://dugi-doc.udg.edu/", assigns a unique and persistent URL to access the document and dataset following the format: http://hdl.handle.net/10256/.

3.1.3. Outline naming conventions used

It has not been decided at M6 yet, but the project dataset identification will likely follow a naming based on the following convention: Data_<WPno>_<serial number of dataset>_<dataset title>. Example: Data_WP2_1_User generated content.

3.1.4. Outline the approach towards search keywords

Data sets have to be findable easily, rapidly and identically. Therefore, standard measures have to be used to identify the data sets. This can include the definition and use of naming conventions, search keywords, version numbers, metadata standards and standard data identifiers. With respect to keywords, the following list is proposed at this stage of the project execution: Energy systems (production, distribution, application), energy collection, conversion and storage, renewable energy, low voltage grid flexibility, advanced power electronics, storage management at grid level, generation and demand forecasting, scheduling, self-healing, monitoring, PMU, cybersecurity, energy business models.

Additionally, it's been agreed to follow the Smart Grid Architecture *Model* (SGAM) to describe the different use cases proposed in the project in WP1, and therefore, all variables naming will be based on this existing model when covered by the model. Moreover, SCADA systems, and WAMS and PMUs, used by EYPESA and CS respectively use tele-control standards IEC-60870-5-104 / EyPESA Profile, 61850 for standards of substation/grid automation and IEC 61970 (Common Information model, CIM). In the next updates of this data management plan, we will revise if all these standards do have or not, a list of keywords and/or meta data.

3.1.5. Outline the approach for clear versioning

The repository will host the final data. It's not a working tool. Thus, there will not be any versioning management for the data used in the project. Moreover, due to the nature of the data (electrical variables such as active power, reactive, power apparent power, voltage, currents, etc.), it is not necessary to consider different versions of them.

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

Each file associated with data will be accompanied with unique specified metadata in order to allow ease of access and re-usability. Standards such as the Dublin Core and Datacite following the guidelines recommended by OpenAire. Standards indicated in section 3.1.4 will be analysed to identify metadata they're using (or not).

3.2. Making data openly accessible

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

In the case of data generated in the laboratory, there is no reason to constrain the open access to the data included. However, the data collected in the real low voltage will have two origins. First EyPESA, as owner of the grid where the pilot will be deployed will provide data collected in its own infrastructure (SCADA, AMI systems). Second instruments and technologies developed in the project will produce other data. The second dataset will be made open whereas the first will be supplied by the company to the consortium under private and confidential conditions. When it is necessary to open this first data, for example to illustrate transformation and dependencies with the second set or to make publications replicable and more relevant, a specific authorization from data owner will be managed.

3.2.2. Specify how the data will be made available

Laboratory data will be made available after publication of corresponding research. Research could be included either in project deliverables or articles. No embargo expected.

Company owned data (EyPESA) will follow two strategies depending on the typology of data:

1) Data needed after publication of a contribution of research: will be opened without constraints.
2) Data needed to exchange information with the platform and stakeholders, without reproducing any specific research result being published: criteria explained in the section 3.2.1 of this document will be followed.

3.2.3. Specify what methods or software tools are needed to access the data. Is documentation about the software needed to access the data included. It's possible to include the relevant software (e.g. In open source code)

Data from experiments, delivered as data tables, will be made available in text files ('csv', 'xml', 'json') easily accessible with any text editor, spreadsheet software or read commands available in any software environment. When necessary a header or configuration file will be included to facilitate the reading. It is also expected to use an API, specifically from smart meters DB should be used the tool Rabbit MS in the case of asynchronous communications. For the SCADA DB it is executed to exchange information from/to folder using text files ('csv', 'xml', 'json') easily accessible as mentioned before.

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

At the moment of this document delivery (M6), there's not a final decision about the repository where the data and publications will be deposited. Universitat de Girona, coordinator of the project, has proposed its institutional repository: https://dugi-doc.udg.edu/ as a candidate; however, other repositories (zenodo) could also be used. It has to be discussed to have an agreement on that. Final decision will be included in future revisions of this document.

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

According to the articles 29.2 and 29.3 of the Grant Agreement, each beneficiary must ensure open access (free of charge online access for any user) to all peer-reviewed scientific publications

relating to its results. Depositing in a research data repository, the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible. Data may be used by third parties under proposed CCBY license taking into account that these data will be used under data protection law according to the agreements achieved whenever necessary.

3.3. 2.3 Making data interoperable

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

In order to facilitate interoperability, we use OAI service, that allows to serve items in XML format for harvesting metadata from other repositories.

OAI server allows to ask for records in different formats (OAI_DC, METS,DIDL, DATACITE -We only allow this format for the openaire-data set).

Besides metadata, METS and DIDL formats offer file download uri and its preservation metadata in order to check file integrity.

OAI_DC and DATACITE formats provide only metadata. The main problem of the OAI_DC format, which is currently the most used standard, is that dublin core metadata element qualifier is lost during harvesting. This may be very confusing, that's the reason why controlled vocabularies for some of these metadata elements have been created.

These vocabularies have a root that provides information on which metadata the values belong to. Specifically, they provide document type, version, access rights and whether they belong to some research program.

For item language we use ISO 639-3 standard. The DATACITE format provides a hierarchical structure that gives related information that adds value to other servers harvesting us (Not specifically DSpace repositories, they don't need to have a Dublin Core metadata system).

Finally, as stated in 3.1.4, Smart Grid Architecture *Model* (SGAM) and standards such as IEC-60870-5-104 / 60870 and standards for substation/grid automation IEC 61970 (Common Information model, CIM, 61850) will be checked in the next updates of this data management plan to analyze if they could have own vocabularies to facilitate interoperability.

3.4. 2.4 Increase data re-use (through clarifying licenses)

3.4.1. Specify how the data will be licensed to permit the widest reuse possible

The documents and files associated to the dataset are proposed to be licensed through an CCBY license. During the project execution, data will be internally available for the consortium partners. It will also be findable and reusable through the final depositing repository (the institutional at UdG has been proposed at M6) and from OpenAire, the latest by the end of the project.

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

The data will remain re-usable after the end of the project by anyone interested in it, with no access or time restrictions.

3.4.3. 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.

Data may be used by third parties under proposed CCBY license taking into account that these data will be used under data protection law according to the agreements achieved whenever necessary.

3.4.4. Describe data quality assurance processes

Since data mining algorithms developed in WP3 will rely on the availability of data specified in section 1, it must be standardized and organized for and easy access and guaranteeing data quality. Application of data preprocessing procedures for cleaning data (missing values, repair of abnormal values, outliers avoided, etc…), and appropriate storage in files are proposed for this purpose.

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

The length of time for which the data will remain re-usable correspond to the time period the repository system will be available, in accordance with the project consortia.

4. Allocation of resources.

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

There are no costs associated to the described mechanisms to make the database FAIR and long term preserved. This is a structural service in the University of Girona not associated to extra costs in the projects the organization is participating. The cost will be covered at the local hosting institute in the context of RESOLVD, as part of the standard network system maintenance.

4.2. Clearly identify responsibilities for data management in your project

The project coordinator, supported by data providers, has the ultimate responsibility for the data management in the project. Moreover, the project coordinator will be the liaison between the data owners and EC.

4.3. Describe costs and potential value of long term preservation

There are no additional costs associated to the described mechanisms to make the database FAIR and long term preserved. For the case of the institutional repository at UdG, this is a structural service in the University of Girona not causing an extra costs in the projects in which the organization is participating.

5. Data security

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

We are differentiating the data being used during the project execution and the final datasets that will be uploaded and made available at the institutional repository at the end of the project.

While the project is running and during the implementation of work, google drive will be used, so a copy of data is automatically performed. This data will not include sensitive information, as is stated in deliverables included in work package 9 (see section 6 in this document). They will be hosted in a Windows 10 enterprise server.

Regarding the institutional repository, full copy is backed up 4 times a year using corresponding exportation and backups systems. In addition, there're several periodical backup on demand, and also before and after main System and applications updates.

6. Ethical aspects to be covered in the context of the ethics review, ethics section of DoA and ethics deliverables.

The ethical aspects are related to the use of personal data and are already addressed in the project in the following deliverables:

- D9.1 (M1) to confirm that the ethical standards and guidelines of Horizon2020 are rigorously applied, regardless of the country in which the research is carried out.
- D9.2 (M1) to provide documents detailing secure data management procedures to guarantee ethics and privacy
- D9.3 (M3) to provide a document with EPESA authorization and conditions of use and manage the data.
- D9.4 4 (to be submitted in M19) detailing how ethics and privacy issues related to data management comply the EU legislation.

7. Other issues

At Spanish level, Law 14/2011 of June 1st, on Science, Technology and Innovation (Article 37 Dissemination in open access) is being considered for data management procedures.

8. Further support in developing your DMP

This DMP has been created with the support tool "Pla de Gestió de Dades de Recerca", http://dmp.csuc.cat

Research Data Management Plan is a development of Digital Curation Center (DCC), adapted by the Consortium of Libraries Universitaries of Catalonia (CSUC). It is based on the open source DMPRoadmap codebase. This institutions work closely with research funders and universities to produce a tool that generates active DMPs and caters for the whole lifecycle of a project, from bid-preparation stage through to completion.

9. Conclusions

This deliverable has provided details about the data management plan envisioned within the RESOLVD project. This is the first version of the DMP delivered after 6 months of project.

This DMP will be updated in parallel with periodic reports and project management plans. The partners providing data (UPC, EYPESA and CS) and the project coordinator serve as references for questions related to data management in RESOLVD.

10. Acronyms and abbreviations

AMI	Advance Metering Infrastructure
API	Application Program Interface
CSUC	Consortium of Libraries Universitaries of Catalonia
CIM	Common Information model
DB	Data Base
DMP	Data Management Plan
DoA	Description of Action
DCC	Digital Curation Center
DSO	Distribution System Operator
FAIR	Findable, Accessible, Interoperable and Reusable
KPI	Key performance Indicator
LV	Low Voltage

MDC	Meter Data Collector
ORD	Open Research Data
PMU	Phasor measurement Unit
RTU	Remote Terminal Units
R+D	Research and Development
SCADA	Supervisory Control And Data Acquisition
w.r.t	with respect to