



# Data management plan V1

## START DATE OF THE PROJECT

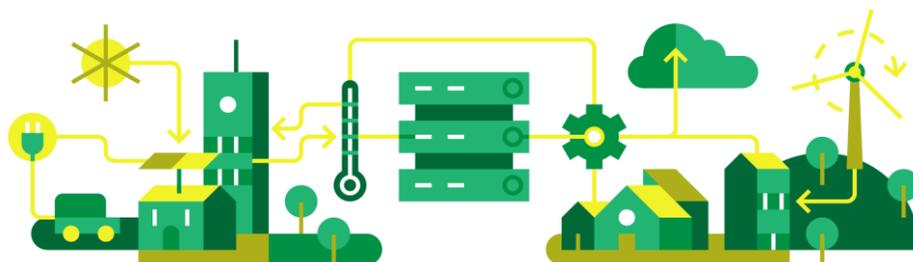
01/10/2017

## DURATION

36 months



The CATALYST project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 768739. Any dissemination of results must indicate that it reflects only the author's view and that the EASME is not responsible for any use that may be made of the information it contains.



<b>PROGRAMME NAME</b>	ENERGY EFFICIENCY CALL 2016-2017
<b>PROGRAMME IDENTIFIER</b>	H2020-EE-2016-2017
<b>TOPIC</b>	Bringing to market more energy efficient and integrated data centres
<b>TOPIC IDENTIFIER</b>	EE-20-2017
<b>TYPE OF ACTION</b>	IA Innovation action
<b>PROJECT NUMBER</b>	768739
<b>PROJECT TITLE</b>	CATALYST
<b>COORDINATOR</b>	ENGINEERING INGEGNERIA INFORMATICA S.p.A. (ENG)
<b>PRINCIPAL CONTRACTORS</b>	SINGULARLOGIC ANONYMI ETAIREIA PLIROFORIAKON SYSTIMATON KAI EFARMOGON PLIROFORIKIS (SiLO), ENEL.SI S.r.l (ENEL), ALLIANDER NV (ALD), STICHTING GREEN IT CONSORTIUM REGIO AMSTERDAM (GIT), SCHUBERG PHILIS BV (SBP), QARNOT COMPUTING (QRN), POWER OPERATIONS LIMITED (POPs), INSTYTUT CHEMII BIOORGANICZNEJ POLSKIEJ AKADEMII NAUK (PSNC), UNIVERSITATEA TEHNICA CLUJ-NAPOCA (TUC)
<b>DOCUMENT REFERENCE</b>	CATALYST.D7.1.PSNC.WP7.v1.2
<b>WORKPACKAGE:</b>	WP7
<b>DELIVERABLE TYPE</b>	Report
<b>AVAILABILITY</b>	Public
<b>DELIVERABLE STATE</b>	Final
<b>DATE OF DELIVERY</b>	30/03/2018
<b>DOCUMENT TITLE</b>	Data management plan V1
<b>AUTHOR(S)</b>	Ariel Oleksiak (PSNC), Marilena Lazzaro (ENG), Vasiliki Georgiadou (GIT)
<b>KEYWORDS</b>	Data management plan; data protection





# Table of Contents

Table of Contents .....	3
List of Tables .....	5
List of Acronyms .....	6
Executive Summary.....	7
1 Introduction .....	8
1.1 Background: Data Management in Horizon 2020.....	8
1.2 Main goals of the CATALYST Data Management Plan .....	8
1.3 Document overview .....	9
2 Data Summary .....	11
2.1 Data sources and owners in CATALYST .....	11
2.1.1 Technical data.....	11
2.1.2 Social data.....	15
3 FAIR Data.....	16
3.1 Making data findable, including provisions for metadata .....	16
3.1.1 Identification of data (naming conventions, versioning) .....	16
3.1.2 Metadata .....	16
3.2 Data sharing – procedures & repositories.....	17
3.3 Making data interoperable .....	18
3.4 Increase data re-use (through clarifying licences).....	19
3.4.1 Licensing .....	19
4 Allocation of resources.....	21
5 Data security and Ethical aspects.....	22
6 Data sets .....	24
6.1 Data set structure.....	24
6.2 Data sets.....	25
6.2.1 Technical data sets .....	26
6.2.2 Social data sets.....	33
7 Conclusions.....	35
8 References .....	36



## List of Figures

FIGURE 1 - PSM PILOT .....	12
FIGURE 2 – PSNC PILOT .....	13
FIGURE 3 - SBP PILOT .....	14
FIGURE 4 - QRN PILOT.....	14
FIGURE 5 - DATA SET SOURCES AND PROCEDURE FOR DATA SHARING DECISIONS .....	18



# List of Tables

TABLE 1 - STRUCTURE OF CATALYST PILOT DATA.....	17
TABLE 2 - DATA SETS TEMPLATE .....	25



## List of Acronyms

<b>DC</b>	Data Centre
<b>DMP</b>	Data Management Plan
<b>DoA</b>	Description of Action
<b>DPIA</b>	Data Protection Impact Assessment
<b>DPO</b>	Data Protection Officer
<b>DSR</b>	Demand Side Response
<b>EU</b>	European Union
<b>GDPR</b>	General Data Protection Regulation
<b>GDC-SG</b>	Green Data Center Stakeholder Group
<b>HPC</b>	High Performance Computing
<b>KPI</b>	Key Performance Indicator
<b>MaaS</b>	Marketplace as a Service
<b>OPC</b>	OLE for Process Control
<b>PSM</b>	Point Saint Martin
<b>PSNC</b>	Poznań Supercomputing and Networking Center
<b>PV</b>	Photovoltaic
<b>RES</b>	Renewable Energy Source
<b>UPS</b>	Uninterruptible Power Supply



## Executive Summary

The deliverable provides an initial version of the CATALYST data management plan (DMP). The document outlines data that will be collected or generated during the CATALYST project and discusses how it will be handled during and after the project lifetime. In the document the first proposal of data management methodology along with data protection measures is proposed. Due to the delivery date at the early stage of the project, the first version of the data management plan concentrates on potential measurement data from CATALYST pilots. Additional data sets based on generated or simulated data (e.g., concerning marketplaces) will be detailed after a final definition of use cases and trials, and described in the final version of DMP in month 18. The final version of this deliverable will also include more detailed plan regarding external data sharing and licensing. This document follows the template provided by the European Commission in the Participant Portal.



# 1 Introduction

In this chapter, we introduce a purpose of building a Data Management Plan by the CATALYST consortium and provide a structure of the document.

## 1.1 Background: Data Management in Horizon 2020

According to the EC Horizon 2020 Work Programme all project proposals submitted to "Research and Innovation actions" and "Innovation actions" have to include a section on research data management which is evaluated under the criterion 'Impact'. Projects participating in the pilot action on open access [1] to research data have to develop a data management plan (DMP) [2] to specify how data will be managed and which data will be open. The DMPs are introduced in the H2020 Work Programme<sup>1</sup>:

“... Horizon 2020 ... 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 purpose of a DMP is to provide an analysis of the main elements of the data management policy that will be used by the applicants with regard to all the data sets that will be generated by the project.

Considering privacy and data protection issues scientific research data should be easily discoverable, accessible, assessable and intelligible, useable beyond the original purpose for which it was collected and interoperable to specific quality standards.

The CATALYST DMP will take into consideration Article 29.3 in the Grant Agreement (Open access to research data). According to the article project participants will have to provide information about data and relevant tools needed for the validation of project results. To this end, data must be stored in data repositories to make it available to third parties who should be able to access, mine, exploit and disseminate the data, free of charge for the user (according to a definition of the open research data).

Importantly, access to parts of data can be limited or denied if this access could cause any risk for the goals of the project. However, in such cases the DMP must explain the reasons for not providing access.

## 1.2 Main goals of the CATALYST Data Management Plan

The CATALYST project takes part in the Open Research Data Pilot. To this end, it defines the Data Management Plan (DMP). This document provides a first version of DMP. The updated and final version is due at M18.

---

<sup>1</sup> The CATALYST project takes part in the Open Research Data Pilot. To this end, it defines the Data Management Plan (DMP). This document provides a first version of DMP.

The main role of the CATALYST DMP is to identify, describe, and propose procedures to manage the data collected and processed by the CATALYST consortium in order to realize project objectives, which are as follows:

- CATALYST will adapt, scale up, validate and deploy an innovative, adaptable and flexible technological and business framework by leveraging on FP7 GEYSER and DOLFIN TRL 4/5 results, with the aim to provide DCs with a set of TRL 6/7 enabling solutions and tools, which:
  - Use and trade the wasted DC heat to lower the energy footprint, reduce DC energy costs and create a new DC income source over longer times;
  - Assess resiliency of energy supply and flexibility, against adverse climatic events or abnormal demand, trading off DC assets energy generation/consumption against local/distributed RES, energy storage and efficiency;
  - Deliver energy flexibility services to the surrounding energy (power and heat) grids ecosystems;
  - Exploit migration of traceable ICT-load between federated DCs, matching the IT load demands with time-varying on-site RES surplus availability or where heat generation is needed (follow the energy approach).
  - Implement novel multi-carrier marketplace mechanisms (in the form of MaaS) to support novel ESCO 2.0 like business models and secure/traceable micro-contracts.
- The CATALYST framework will be adaptable to a broad variety of DCs categories, ranging from different DCs types (co-location, enterprise, HPC DCs), to different geographical location, to different architectures (large centralized versus fully decentralized micro-DCs) and energy efficiency orientation.

To realise these objectives the CATALYST project will collect information on DC energy (electricity, heat and cool) consumption and generation as well as other resource usage, for example water consumption and IT systems utilization. Based on this information CATALYST will prepare a detailed Data Management Plan (updated version of this document) accommodating all necessary measures for proper data handling. This first version of the DMP starts to outline how data collected or generated by the CATALYST project will be organised, stored and shared.

The aim is to gather data valuable for technological and scientific evaluation of the project achievements respecting privacy and legislation (e.g. as mandated by GDPR EU Directive 2016/680 and Regulation 2016/679).

Those project data which are allowed for open access will be anonymised and stored in repositories including publishing to well-known open platforms (e.g. OpenAIRE [3]).

## 1.3 Document overview

This report is a first version of the DMP which will be updated as the project progresses in M18 when all data or potential uses will be more clear. Indeed, the DMP can be seen as a living document which will be officially released as deliverable D7.1 at M6 and as deliverable D7.2 at M18.

The reminder of this report is as follow:

- Section 2 **Errore. L'origine riferimento non è stata trovata.** contains a general description of data collected and used by the CATALYST project.

- Section 3 describes how these data can be identified, shared and made interoperable.
- Section 4 explains the way of financing the DMP by the project.
- Section 5 contains security and ethics considerations along with references to other regulatory frameworks.
- Section 6 introduces a template to describe CATALYST data sets along with first definitions of data sets.



## 2 Data Summary

In this chapter, a general description of data collected and used by the CATALYST project is introduced. The description includes information about data sources and owners in CATALYST, basic data characteristics, and data sets use. More detailed description of data including data sets identified until now are provided in Section 6.

### 2.1 Data sources and owners in CATALYST

In CATALYST, data used comes from 2 main types of sources:

- Technical data related to pilots and trials executed in the project
- Social Data related to people interacting with project

#### 2.1.1 Technical data

Technical data collected and processed within the project will come mostly from CATALYST pilots. That means data collected both from real data centres and separated testbeds accompanied by data from surrounding systems such as renewable energy sources, energy management systems, buildings, etc. As pilots are already defined in the project preliminary descriptions of the relevant data sets have been included into this deliverable. In the second version of the Data Management Plan it will be detailed as at that time (M18) specifications of all trials should be done.

Apart from the data that come directly from project pilots (and partners) we anticipate that other sources related to markets considered within the project may appear. Those data sets can be either generated by simulation tools or taken from external source if available. As they have not been defined yet their definition will be added in the second version of the Data Management Plan, deliverable D7.2.

The purpose of these data sets collection and processing is first of all to test and verify CATALYST platform and solutions in several different environments. Data coming from real environments will be used to fill CATALYST data model and based on it perform optimisations to verify potential gains. Various data sets will be used to validate the project approach in different settings. Re-using these data sets will help in improvements of the project outcomes. Those data sets that can be shared and re-used will allow others to evaluate requirements and benefits of the CATALYST platform.

The pilots that are the main source (and owners) of data collected and process within CATALYST is shortly summarised below.

#### **Pilot 1: ENG PSM – Co-location centre**

This data centre of Engineering is located in Pont Saint Martin. It is a colocation data centre. The trials to be defined in the PSM DC are as follows:

- Energy savings from free-air cooling and from cold water from the aquifer (geothermal energy)
- Heating of DC offices based on surplus heat from DC

- Energy Demand Side Management through energy heat reuse practices and backup generators usage
- DR programs to reduce energy consumption (specific collaboration agreement is already in place)
- Load migration between the Green PSM, PSNC DC and legacy SiLO DC will be tested to a limited extent, due to mission critical business operations.

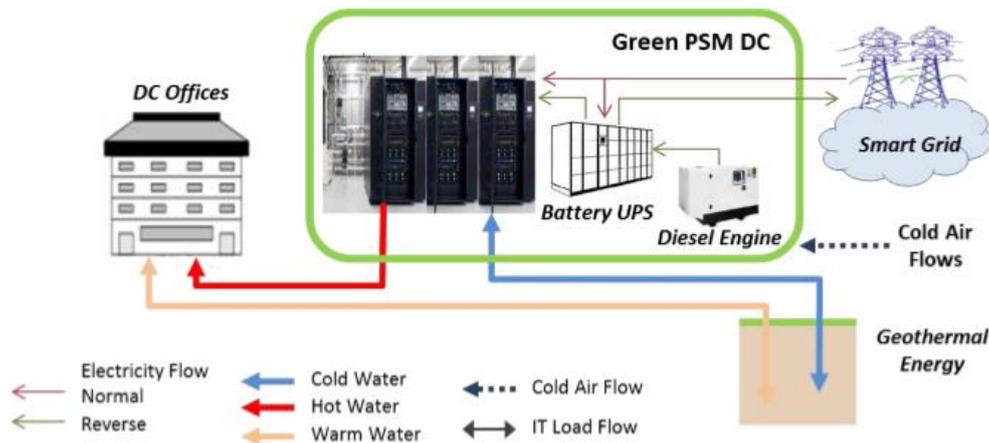


Figure 1 - PSM Pilot

## Pilot 2: PSNC – HPC

This data centre of PSNC is mainly a High Performance Computing (HPC) data centre. PSNC also provides a testbed (micro data centre) with possibility of connections to renewable energy source (photovoltaic). The trials to be defined in the PSNC pilot infrastructure are as follows:

- Liquid-based DC cooling
- Transfer of heat from liquid loop to the water loop to provide heat to the DC offices (Simulate the interconnection of the water loop to provide district heating to the University campus)
- Interconnect RES (PV cells) and battery based UPS supported to offer electricity & flexibility services
- Host and migrate IT load from DCs (while preserving SLAs) of the same administrative domain and from different administrative domains (PSM DC, SiLO DC and QRN distributed DC)

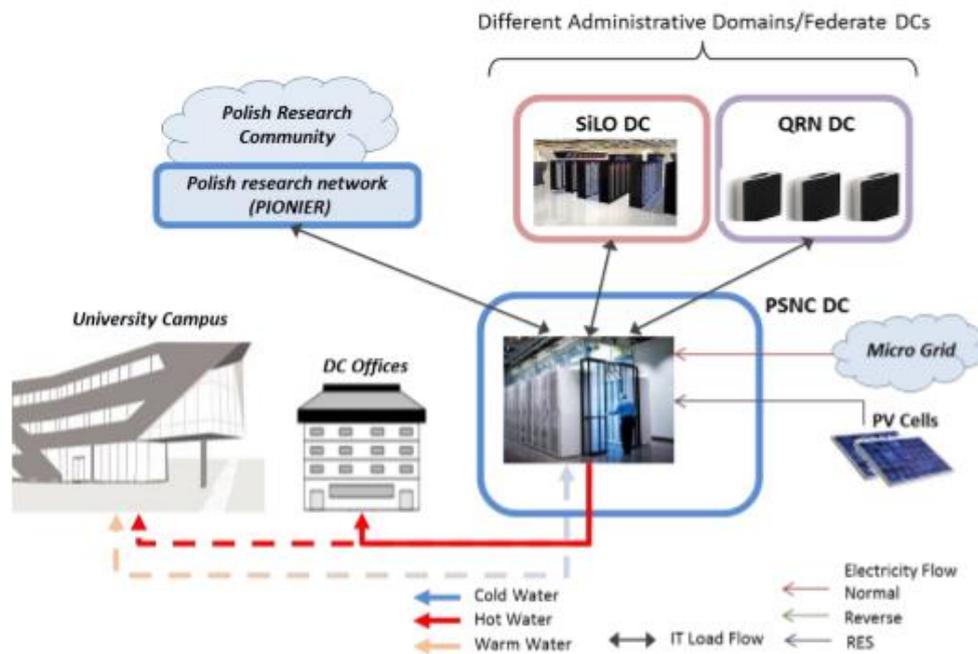


Figure 2 – PSNC Pilot

### Pilot 3: Schuberg Philis – Co-location centre

This data centre is a SBP colocation data centre. The trials to be defined in the SBP DC are as follows:

- Reuse the warm air to provide heat to the DC offices
- Reuse the warm air to pre-heat the diesel UPS engines
- Provide a comparison between battery based and dynamic rotating UPS
- Reuse the cold drinking water to provide energy-free (passive cooling) water cooling
- Collaborate with ALD in electricity DR programs to offer flexibility/ancillary services and reduce energy consumption and provide electricity back for Smart Grid stabilization from dynamic rotating UPS
- Host IT load from the QRN fully distributed DCs if needed (especially during summer period)

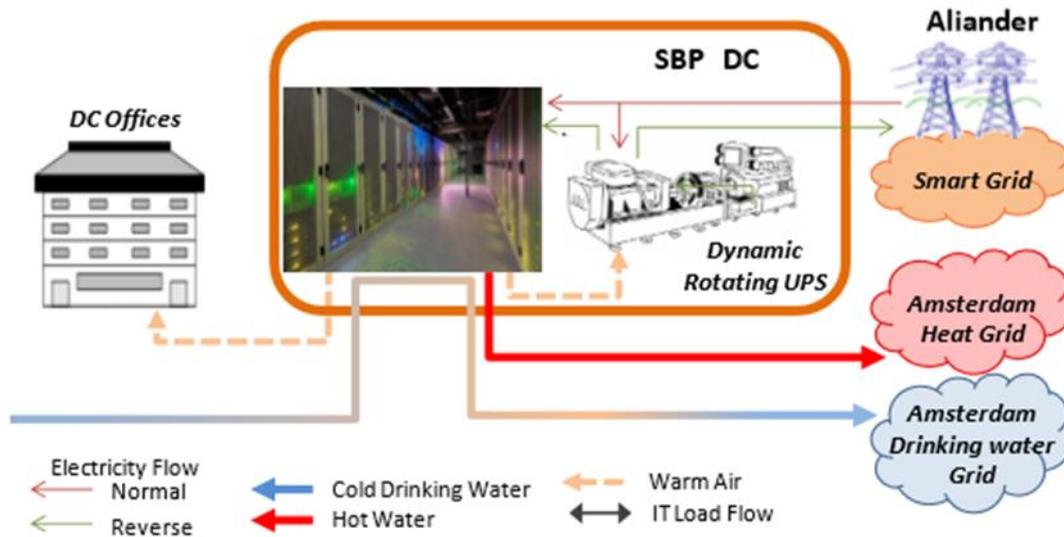


Figure 3 - SBP Pilot

#### Pilot 4: QRN «Need for heat »- Distributed HPC

QRN has a distributed data centre model. Therefore QRN pilot consists of multiple servers integrated with radiators deployed in a building. The trials to be defined in the QRN DC are as follows:

- Stabilize the grid by forecasting heat needs and capacity
- Balance IT loads by migrating VCs within and outside the virtual DC
- Implement DR (electricity and IT load) for grid flexibility
- Measure heat response and adequacy to temperature command

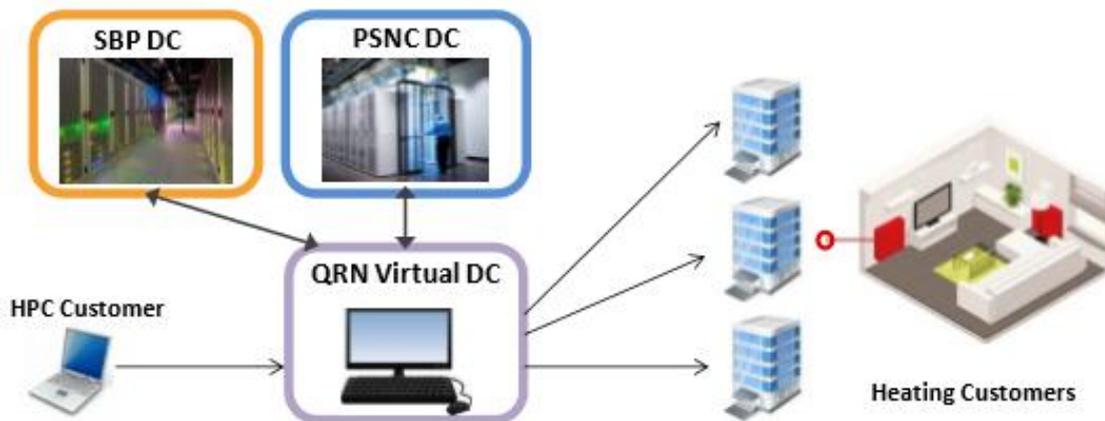


Figure 4 - QRN Pilot

As can be easily seen, Pilots represent different types of DCs, surrounding ecosystems, and interactions with external entities. Consequently, pilots are sources of various types of data with different sizes, level of protection, and purpose.

Generally, data types and formats contains time series of numbers related to operation of data centres. The size of data sets can be in the order of magnitude from dozens of MB to even hundreds of GB depending on the trial type and length.

Details and goals of specific data sets are defined in data sets descriptions in Section 6

## 2.1.2 Social data

In addition to the technical work on trials with the use of software developed within the project and applied to CATALYST pilots, it is foreseen that data sets with information about people interacting with the project will be also collected. For example, data about Green Data Centre Stakeholder Group (GDC-SG) members, possibly with their opinions and knowledge, will be processed within the project. Therefore, initial data set for this kind of data has been also defined in this deliverable. Depending on the involvement of GDC-SG members as well as input provided more detailed data sets will be provided in the second version of DMP, deliverable D7.2. It will be also complemented by any other social data that can be collected.

The purpose of creating these data sets is to receive feedback from relevant stakeholders of the energy flexibility ecosystems based on data centres. In addition to the internal use for steering of project development the data sets related to GDC-SG may be used as interesting source of information about future of data centres and as a validation of applicability of the CATALYST concepts and results.

Generally, data will consists of personal data, opinions, minutes, questionnaires. Sizes of these data sets will be relatively small – in the order of magnitude of several MB.

Frist definition of a data set related to GDC-SG can be found in Section 6.

## 3 FAIR Data

### 3.1 Making data findable, including provisions for metadata

#### 3.1.1 Identification of data (naming conventions, versioning)

Data sets will be identified based on a simple taxonomy (pilot/external/simulated/other, type of DC, part of DC: whole, tested, cooling subsystem, etc.):

- Type of infrastructure/system (colocation/enterprise/cloud/ HPC data centre, renewable energy source, smart grid, etc.),
- Purpose of infrastructure (testbed, real production infrastructure or mix),
- Identifier: [TrialID/PilotID/InfrastructureID]

Additionally, the data sets will be identified by time, in which data was collected.

#### 3.1.2 Metadata

In addition, to identification of a data set by the identifier and time, other structured information will be provided in data sets. First information about pilots will be collected according to the structure defined within the project.

Area	Structure
Trial site type (Testbed / DC part / whole DC)	Data centre type <ul style="list-style-type: none"> <li>• e.g. colocation, cloud, HPC (High Performance Computing), enterprise, distributed</li> </ul>
	CATALYST testbed scope <ul style="list-style-type: none"> <li>• relation to main DC, type of resources, size and power usage</li> </ul>
Relevant technologies / specific parts	IT/software: workload types, applied middleware and tools <ul style="list-style-type: none"> <li>• e.g. virtual machines, batch jobs - queuing system, containers, services, other</li> </ul>
	Cooling type <ul style="list-style-type: none"> <li>• e.g., traditional air-based with chillers, Free cooling, DLC, other</li> </ul>
	Renewable Energy Sources <ul style="list-style-type: none"> <li>• e.g., PV, wind, geothermal) and their integration (e.g. on/off-grid, energy storage</li> </ul>
	Heat re-use
	Smart grid interaction <ul style="list-style-type: none"> <li>• e.g., DC as prosumer, DSR (Demand Side Response) - which type?, other</li> </ul>
	Other
Measured data	Power/energy - which granularity <ul style="list-style-type: none"> <li>• e.g. whole DC, cooling, per rack/server, etc.</li> </ul>
	IT <ul style="list-style-type: none"> <li>• e.g. utilization, server states, services, etc.</li> </ul>
	Environmental data <ul style="list-style-type: none"> <li>• e.g., temperatures, humidity, flow, etc.</li> </ul>

	Cooling system <ul style="list-style-type: none"> <li>e.g. delta T, inlet/outlet temp, water temp, flow</li> </ul>
	RES production and use <ul style="list-style-type: none"> <li>e.g. energy produced, consumed, other parameters</li> </ul>
	Heat re-used <ul style="list-style-type: none"> <li>e.g. amount of heat</li> </ul>
	Other
Metrics	PUE <ul style="list-style-type: none"> <li>e.g. Level 1-3, Partial</li> </ul>
	Energy to solution <ul style="list-style-type: none"> <li>e.g. energy per job</li> </ul>
	RES related metrics <ul style="list-style-type: none"> <li>e.g. ERE</li> </ul>
	Cooling efficiency metrics <ul style="list-style-type: none"> <li>e.g. COP</li> </ul>
	Reliability metrics <ul style="list-style-type: none"> <li>e.g. availability, MTBF</li> </ul>
	Performance metrics <ul style="list-style-type: none"> <li>e.g. throughput</li> </ul>
	Other
Possible actions/strategies	IT equipment management <ul style="list-style-type: none"> <li>e.g. CPU throttling, servers on/off, power capping</li> </ul>
	Software management <ul style="list-style-type: none"> <li>e.g. VM consolidation/suspend, batch scheduling, application/services management, etc.</li> </ul>
	Cooling control <ul style="list-style-type: none"> <li>e.g. room/coolant temp adjustments, etc.</li> </ul>
	Additional DSR actions <ul style="list-style-type: none"> <li>e.g. UPS or diesel engines for Smart Grid stabilisation</li> </ul>
	Other actions

Table 1 - Structure of CATALYST pilot data

In further stages of the project it will be synchronised with the CATALYST data format.

## 3.2 Data sharing – procedures & repositories

Data will be shared in the following way:

- Repositories of testbeds such as <http://labee.psnk.pl> (at PSNC)
- Storage systems for monitoring data in data centres
- External repositories such as OpenAIRE

To this end, restrictions on access must be defined for each data sets (see Section 6). To control this process a Data Access Committee will be established consisting of project coordinator, data management officer and data owner.

The sources of data sets in CATALYST along with a procedure for data sharing decisions is illustrated in Figure 5.

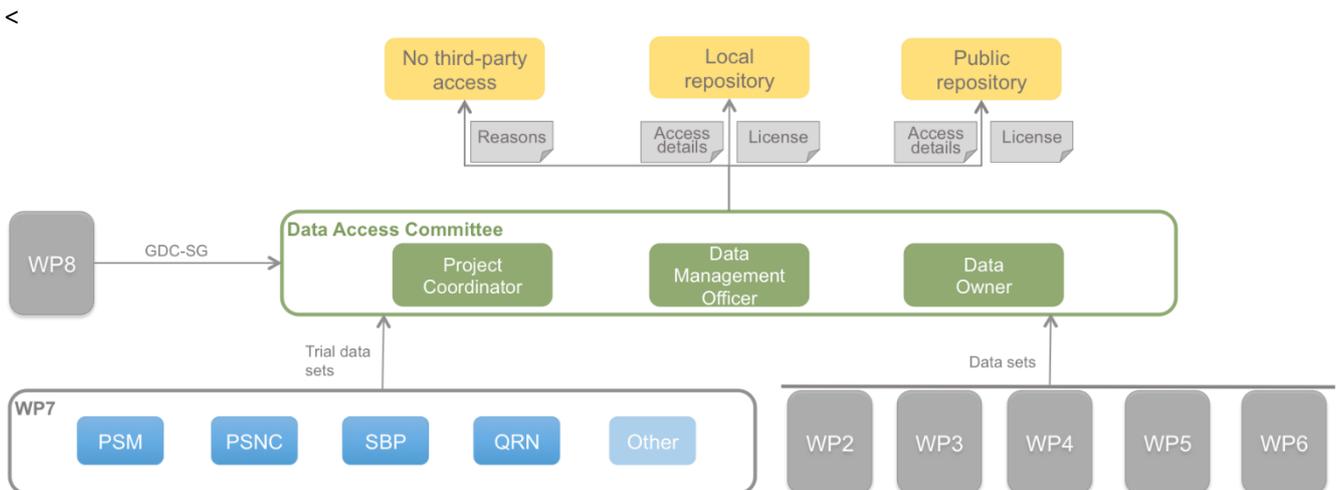


Figure 5 - Data set sources and procedure for data sharing decisions

Data sets will be mostly generated by pilots within WP7, however, other sources related, for instance, to marketplaces or any other data generated/used by other work packages are also possible as indicated in the diagram. When data set is defined and is going to be collected and/or used in the project the data owner or WP leader must inform the Data Management Officer. The Data Access Committee decides whether and how this data set will be shared. If there are restrictions due to internal policies of data owners decision about denying access can be taken. The Data Access Committee may also decide to pre-process data before sharing, for example, performing anonymization or restricting access to only parts of data. For shared data license and access details will be defined (e.g. location, identifiers). Depending on the volume, usefulness and time when information is outdated specific sets can be accessible from local repositories or public open repositories such as OpenAIRE.

### 3.3 Making data interoperable

If possible, CATALYST will follow existing good practises, standards and data formats to make data interoperable and ensure compliance.

Plan towards interoperability will include:

- Application of recognized CATALYST good practices, e.g.:
- Generic format used by DOLFIN to store measurement data – flexible enough for various data description
- Data model based on GEYSER experiences to describe the data centre along with its ecosystem (interfaces to marketplaces, etc.) for the CATALYST optimiser
- Taxonomy for describing DC applied in CATALYST – based on existing classifications of DC, their subsystems, etc. (as presented in Section 3.1.2)
- Taking into consideration a compliance with relevant standards and regulations when establishing measurement procedures, data naming, etc.,

- Applying Green Grid PUE monitoring guidelines [4] when performing energy measurements and calculating PUE (standard PUE – EN50600-4-2)
- If calculating the Renewable Energy Factor (REF) referring to standard EN50600-4-3 [5]
- Checking if applied strategies are related to best practices defined in the EU Code of Conduct (and CLC/TR 50600-99-1 standard to which these practices have been incorporated) [6]
- Referring to ASHRAE TC 9.9 2015 for thermal guidelines and server types [6]
- Looking at other EN50600 standard series, e.g. EN50600-4-4 and EN50600-4-5 if KPIs such as IT Equipment Energy Efficiency and Energy Utilisation for Servers must be monitored [5]

## 3.4 Increase data re-use (through clarifying licences)

Data reusability implies a deep knowledge of both content and provenance of data sets as well as an accurate understanding of their license. The aim of D7.1 is to provide information about the how the project aims to promote the widest reuse of data. This report is an initial version of the DMP providing a first overview of data collected/generated into the project. A further version will be released at M18 when the project progresses will allow to create a more detailed data sets. As a consequence, the next version of DMP, deliverable D7.2, will pay particular attention to enabling reuse of data sets and will state which data can be shared also beyond scope and lifetime of CATALYST.

The following subsection introduces some preliminary information about the CATALYST approach to licensing.

### 3.4.1 Licensing

The data management plan establishes how licensing will be addressed in the project lifetime according to the identified data set. At this stage of the project, it could be useful, to start classifying the project data sets and provide the initial positioning of the project in reference to licensing. In the frame of the project, three main categories of data are identified as follows:

- Project data sets for dissemination (e.g. Social data sets): this kind of data will be open access by default in order to validate scientific publications.
- Data sets produced by work packages: this kind of data will be reused in accordance with the specific WP policy. For example some data created and stored during the project can be for internal management and communication within the consortium. Some other can be shared outside the project and useable by third parties.
- Technical data sets: those data are about the Pilots and are mainly about pilots measurement collected within the project timeframe. Simulation data are considered as well. Currently, it is hard to define exactly what license can be applied to the pilots-related data sets. However some assumptions can be done as follows:
  - Data sets that enclose personal data cannot be shared as-is for privacy reasons. This is the case of QRN DC which handle multiple radiators deployed in residential/office buildings. Even if these data cannot be tied to one person or household, they have to be treated cautiously; the QRN data set will be shared only within the CATALYST project members upon request.
  - Data sets that require aggregation or anonymizing for security or commercial reasons prior to release. This is the case of PSM DC: row data can be shared only for the objective of the trial and

in the scope of the project; otherwise aggregated data, the final result of the CATALYST solution exploited in the PSM Pilot, can be shared not only for the objective of the trial.

- Data sets that can be shared without any restriction: this is the case of PSNC Micro data centre testbed and data from SBP DC
- Data sets that cannot be shared outside the project: this is the case of PSNC DC related to the High Performance Computing centre.

A further public version of the Data Management Plan will be released at M18 as deliverable D7.2. The DPM will be properly updated and it will consider in detail:

- the refined version of data sets already reported in D7.1;
- Identification of new data sets of the project;
- what licenses and/or restrictions are applicable to the data sets;
- which data sets are useable by third parties, in particular after the end of the project;
- if the re-use of some data sets is restricted explaining why.

## 4 Allocation of resources

Allocation of resources is about costs for making research data findable, accessible, interoperable and re-usable (FAIR) in the project.

These costs can be different according to the specific data sets that is considered.

Regarding the provision of DMP, the related costs involves WP7 - *Trials & Performance Validation* since Task 7.1 - *Specification of trials and evaluation methodology* is responsible for the preparation of data management plan. These costs are covered by budget associated to WP7.

PSNC as WP7 leader is responsible along with other WP7 (especially T7.1 - ENG) partners for the overall data management process.

Resources can be also related to the long term preservation of data. How those data will be kept and for how long will be discussed in D7.2 in accordance with the updated and refined version of CATALYST data sets. However it can be assumed that storage of pilot-related data sets and potential access beyond the project lifetime will be covered by partner formally responsible of CATALYST pilots.

## 5 Data security and Ethical aspects

### Data security

Data security is about the provision of data protection in the project lifetime. The aim of DMP is to provide information about data recovery, secure storage and transfer of sensitive of data as well.

At this stage of the project, it can be stated that pilots-related data sets will be stored in pilot DCs where there is high level of security and redundancy to prevent failures and data losses.

Certified repositories will be taken into account for project data that has to be curated and preserved for long time. Some repository like OpenAIRE will be considered to share data outside the project.

More detailed information will be provided in D7.2 according to the specific data set that is identified and discovered.

### Ethical aspects

Ethical aspects mainly concern personal data. If this data has to be shared, this aspect has to be taken into account. Indeed, as stated in chapter 5 of CATALYST Description of Action (DoA), the Consortium is aware of EU and national legislation and policies referring to protecting personal data and privacy, especially in the context of smart grid. The CATALYST consortium is committed to taking all necessary measures to ensure that all project activities comply with the European Chart of Fundamental rights and all data protection relevant EU regulations, soft – law, standardisation and policy initiatives.

To ensure appropriate protection of data sets CATALYST will be tracing, checking relevance for CATALYST data sets, and if needed adopting at least the following procedures/regulations:

- GDPR (General Data Protection Regulation) for data protection
- Data protection regulation framework in smart energy grids
- The Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data
- The Directive (EU) 2016/680 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences
- Ad-hoc frameworks on data privacy in smart grids, including Data Protection Impact Assessment (DPIA).
- Data Protection Impact Assessment Template (supported by Commission Recommendation 2014/724/EU)

Moreover, the project will assure the respect of personal data protection through a dedicated Data Protection Officer (DPO). It works at project level in order to overview that an appropriate data management plan is developed and used to protect the privacy of data. As reported in the CATALYST DoA, section 3.2.1.1 Project Management Structure, the role of the CATALYST DPO will be compliant to the GDPR (EU 2016/679, EU 2016/680). The CATALYST DPO is Dr. Ariel Oleksiak (PSNC).

Focusing on technical data sets collected so far, in some cases the pilot data sets can contain “personal data” like for the QRN pilot where data come from owners of heaters. Those data will be analysed to ensure

proper anonymization. A proper questionnaire, dealing with personal data, will be distributed, if needed, to the owners/managers of residential/office buildings; the informed consent for data sharing and long term preservation will be considered as well. Indeed, it is important to inform owners that data from their heaters could be processed within CATALYST project.

Regarding Social data sets, potential questionnaires to collect data from GDC-SG members may contain personal data and will include information about goals and rules related to processing these data.

There are no other ethical issues identified so far with respect to the other CATALYST data sets. More details will be provided in the updated version of DMP at M18.

## 6 Data sets

This section contains a first definition of a structure used to define data sets as well as initial data sets foreseen to be created and used in the CATALYST project.

### 6.1 Data set structure

The template used to describe data sets has been based on the guidelines from EC enhanced by (i) specific parameters related to CATALYST needs (such as Trial ID) and (ii) additional guidelines related to given parts (e.g. type of data centre in the case of CATALYST DC pilots).

The template to describe data sets is given below:

Project name	CATALYST
ID	768739
Project coordinator	Massimo Bertoncini
Project description	Converting DCs into Energy Flexibility Ecosystems
Funder	H2020-EE-2016-2017
Principal researcher	Who is the researcher responsible for this data set?
Collecting period	Timeframe within which data were collected
Trial ID	Identifier of the CATALYST trial
Data set reference and name	<p>What data are collected and how are they named?</p> <ul style="list-style-type: none"> <li>Type of infrastructure/system (colocation/enterprise/cloud/ HPC data centre, renewable energy source, smart grid, etc.),</li> <li>Purpose of infrastructure (testbed, real production infrastructure or mix),</li> <li>Identifier: [TrialID/PilotID/InfrastructureID]</li> </ul>
Data set description	<p>Description of the data, the origin, nature and scale and for what purpose they were generated.</p> <p>Type of data, e.g.</p> <ul style="list-style-type: none"> <li>Power/energy measurements: of servers, cooling systems, other devices</li> <li>Energy production and supply: RES production, heat re-used,</li> <li>Environmental data: temperatures in server rooms, humidity, flow, etc.</li> <li>IT monitoring: utilization of servers, VM management,</li> <li>Energy market: prices of energy</li> <li>Other</li> </ul> <p>Source</p> <ul style="list-style-type: none"> <li>Origin: laboratory/infrastructure measurements, experiments, simulations, etc.</li> <li>Specific subset of infrastructure</li> </ul> <p>Volume</p> <ul style="list-style-type: none"> <li>Total volume of data, number of files, etc.</li> </ul> <p>Data and file formats</p> <ul style="list-style-type: none"> <li>Format of data</li> </ul>

	<ul style="list-style-type: none"> <li>Standards, open/common formats used</li> </ul>
Standards and metadata	<p>Are there any suitable standards and what metadata will be created?</p> <p>Metadata to identify data set</p> <ul style="list-style-type: none"> <li>Identification of trial, pilot and specific part of infrastructure (or other systems and sources used) for which this data set has been prepared and/or used</li> </ul> <p>More advanced metadata to (possibly to be synchronised with a CATALYST data model)</p> <ul style="list-style-type: none"> <li>help to understand and interpret data</li> <li>provide details about experiment setup (who, when, in which conditions, tools, versions, etc.)</li> <li>help identify and discover new data</li> </ul> <p>Any community standards used to enable interoperability</p>
Data sharing	<p>Describe if and how the data will be shared (access, procedure, embargo periods, technical mechanisms, necessary software and tools, repositories). In case the data cannot be shared describe the reason.</p> <p>Which data will be shared?</p> <ul style="list-style-type: none"> <li>Final result?</li> <li>Intermediate data?</li> </ul> <p>Data protection</p> <ul style="list-style-type: none"> <li>Who will have access?</li> <li>Is there a need to preprocess the data, e.g. anonymize?</li> <li>What are constraints and reasons of limited access if this is the case</li> </ul>
Archiving and preservation	<p>Where and how long will the data be preserved? What is the approximated end volume, what are the associated costs and how will they be covered?</p> <p>Which data should/needs to be preserved?</p> <ul style="list-style-type: none"> <li>What has to be kept e.g. data underlying publications?</li> <li>What can't be recreated e.g. environmental recordings?</li> <li>What is potentially useful to others?</li> </ul> <p>For how long?</p> <p>What is the final size of data?</p> <p>Where it can be stored?</p> <p>What is the cost and who will pay for it?</p>

Table 2 - Data sets template

## 6.2 Data sets

As explained in Section 2, data sets defined so far consists of data sets corresponding to specific pilots available in CATALYST (4 technical data sets in Section 6.2.1) and a data set related to people and their

opinions (a data set for data from GDC-SG in Section 6.2.2). Additional data sets based on generated or simulated data (e.g. concerning marketplaces) will be detailed after a final definition of use cases and trials is done and described in the final version of DMP in month 18.

## 6.2.1 Technical data sets

### PSM-ENG Pilot

PSM-ENG pilot provides a colocation data centre Point Saint Martin belonging to the Engineering group.

<b>Project name</b>	<b>CATALYST</b>
ID	768739
Project coordinator	Massimo Bertoncini
Project description	Converting DCs into Energy Flexibility Ecosystems
Funder	H2020-EE-2016-2017
Principal researcher	Marilena Lazzaro
Collecting period	The data set will be collected in the timeframe within the application of use cases in the PSM Pilot: M16 (January 2019) - M36 (September 2020).
Trial ID	TBD based on definition of use cases with the use of PSM Pilot
Data set reference and name	<p>[TrialID/PSM]</p> <ul style="list-style-type: none"> <li>• colocation data centre</li> <li>• Mix: measurement data from the real production DC infrastructure; only data relevant for the CATALYST trials will be part of the testbed.</li> <li>• Identifier: TrialID/PSM/ColocationDC</li> </ul>
Data set description	<p>Measurement data from both facilities and IT</p> <p>Type of data:</p> <ul style="list-style-type: none"> <li>• Power/energy measurements: Electrical consumption of cooling devices, Main incomer, Server room incomer, UPS.</li> <li>• Energy production and supply: geothermal system is used for cooling the DC (bunkers and offices), heat recovery system to supply heat to office buildings.</li> <li>• Environmental data: temperatures in server rooms, Chilled air/water flow rate, Relative/Absolute humidity of the server room, Ambient air temperature.</li> <li>• IT monitoring: CPU utilization (%) of servers, Storage utilisation (%), Network utilisation (%)</li> </ul> <p>Source</p> <ul style="list-style-type: none"> <li>• Specific subset of DC infrastructure measurements</li> </ul> <p>Volume</p> <ul style="list-style-type: none"> <li>• From tens of MB to GB depending on the time experimentation will last and on the data that will be necessary for the trials</li> </ul> <p>Data and file formats</p>

	<ul style="list-style-type: none"> <li>• Format of data monitored through the Honeywell OPC Server Data Access meets the specification of the OPC Data Access Standard version 1.0a or 2.0. This data are the Power/energy measurements and the Environmental data.</li> <li>• Format of IT monitoring data depends on the specific tool that is used like NAGIOS for Storage utilisation (%) and ORION for Network utilisation (%).</li> </ul>
Standards and metadata	Data is identified by a related timeframe. Within CATALYST Trial ID will be used to identify data sets. Other metadata to be discussed (possibly to be synchronised with a CATALYST data model).
Data sharing	Row Data from PSM DC can be shared only for the trial's objective and in the scope of the project. How this will be done is under evaluation according to the specific use cases to be demonstrated. The limited access to this data is due to the internal policy of PSM DC. Aggregated Data, the final result of the PSM Pilot, can be shared not only for the objective of the trial.
Archiving and preservation	Currently the PSM DC allows the data archiving according to the specific contract agreed upon the customer. This data are mainly related to alarm events and are stored as .data files. Regarding the data set described in this template, if historical data are needed for the purpose of the trial, the archiving mechanism already in place in the DC will be properly extended to the PSM pilot. Data will be preserved in the DC during the Project and for a period of 1 year after the end of the Project. The associated cost will be covered by ENG.

## PSNC Pilot

PSNC pilot consists of several parts. First of all, Poznań Supercomputing and Networking Center operates High Performance Computing (HPC) Center mostly addressed (not only though) to scientific users. In addition to the operational DC PSNC provide access a testbed of the energy-efficient technologies laboratory. Data will be complemented by detailed monitoring of the photovoltaic system, which can be connected to the testbed.

## Micro data centre testbed

<b>Project name</b>	<b>CATALYST</b>
ID	768739
Project coordinator	Massimo Bertocini
Project description	Converting DCs into Energy Flexibility Ecosystems
Funder	H2020-EE-2016-2017
Principal researcher	Ariel Oleksiak
Collecting period	TBD
Trial ID	TBD based on definition of use cases with the use of PSNC Pilot
Data set reference and name	[TrialID/PSNC/MicroDC] <ul style="list-style-type: none"> <li>• PSNC micro data centre laboratory</li> </ul>

	<ul style="list-style-type: none"> <li>Laboratory/testbed used in CATALYST trials</li> </ul>
Data set description	<p>Measurement data from both facilities and IT. Data from around 20-100 server nodes with power usage in order of 3-10kW. Servers can be partially powered by energy produced by photovoltaic system.</p> <p>Type of data:</p> <ul style="list-style-type: none"> <li>Power usage of servers and cooling;</li> <li>Utilisation of servers;</li> <li>Temperature of servers, room and coolant (for direct liquid cooling);</li> <li>if needed other statistics related to workload execution (e.g. OpenStack, SLURM)</li> </ul> <p>Source</p> <ul style="list-style-type: none"> <li>Micro data centre laboratory – a testbed at which experiments related to CATALYST trials can be executed</li> </ul> <p>Volume</p> <ul style="list-style-type: none"> <li>Currently the size of data is in the order of 30GB per one month and more than 70GB per year (older data is stored with lower sampling rate)</li> </ul> <p>Data and file formats</p> <ul style="list-style-type: none"> <li>Mostly key-value pairs for various parameters. Data can be easily translated into the CATALYST format (partially done for monitoring in DOLFIN).</li> </ul> <p>Purpose</p> <ul style="list-style-type: none"> <li>Experiments with data centres, analysis of trial results</li> </ul>
Standards and metadata	<p>Data is identified by a related timeframe. Within CATALYST Trial ID will be used to identify data sets.</p> <p>Other metadata to be discussed (possibly to be synchronised with a CATALYST data model).</p>
Data sharing	<p>Data from PSNC micro data centre can be shared. Currently are stored in file-based database at PSNC. Intermediate detailed data can be stored in relational data bases.</p> <p>Data is partially available to registered users via <a href="https://labee.man.poznan.pl">https://labee.man.poznan.pl</a></p>
Archiving and preservation	<p>Measurements from micro data centre are currently stored at PSNC for at least period of 1 year. The total volume collected during 1 year currently (using RRD database that automatically decrease size of older data) is equal to 77GB. If the whole high frequency data is stored then, depending on trials timeframe, the total volume can reach up to 0.5TB. This data set can demonstrate details of a trial in which certain results were achieved.</p> <p>These data can be stored at PSNC.</p>

### Photovoltaic system

Project name	CATALYST
ID	768739



Project coordinator	Massimo Bertoncini
Project description	Converting DCs into Energy Flexibility Ecosystems
Funder	H2020-EE-2016-2017
Principal researcher	Ariel Oleksiak
Collecting period	TBD
Trial ID	TBD based on definition of use cases with the use of PSNC Pilot
Data set reference and name	[TrialID/PSNC/RES-PV] <ul style="list-style-type: none"> <li>PSNC photovoltaic system energy production</li> </ul>
Data set description	<p>Monitored energy generation of the PSNC photovoltaic system. The data come from 80 solar modules that can generate up to 20kW in peak and energy storage (lead-acid batteries of 75 kWh capacity).</p> <p>Type of data:</p> <ul style="list-style-type: none"> <li>Values of energy produced in time</li> <li>Other parameters: P, S, Q x4, day yield, grid frequency, grid voltage x3, total yield</li> <li>Data collected every 5s</li> </ul> <p>Source</p> <ul style="list-style-type: none"> <li>PSNC photovoltaic system (inverters, batteries)</li> </ul> <p>Volume</p> <ul style="list-style-type: none"> <li>Currently the size of data is in the order of 25MB per one month and around 50MB per year (older data is stored with lower sampling rate) but may be increased up to 600MB if more data monitored and the whole data stored</li> </ul> <p>Data and file formats</p> <ul style="list-style-type: none"> <li>Mostly key-value pairs of time and power (other parameters also possible). Data can be easily translated into the CATALYST format (partially done for monitoring in DOLFIN).</li> </ul> <p>Purpose</p> <ul style="list-style-type: none"> <li>Analysis of trials with on-site renewable energy use</li> </ul>
Standards and metadata	Data is identified by a related timeframe. Within CATALYST Trial ID will be used to identify data sets. Other metadata to be discussed (possibly to be synchronised with a CATALYST data model).
Data sharing	Data from PSNC photovoltaic system can be shared on request. Currently are stored in file-based and relational data bases at PSNC.
Archiving and preservation	Measurements from photovoltaic system are currently stored at PSNC for at least period of 1 year. Currently, the total volume collected during 1 year currently (using RRD database that automatically decrease size of older data) is equal to 50MB. If the whole high frequency data is stored then, depending on trials timeframe, depending on trials timeframe the total volume can reach 600MB. This data set can demonstrate details of a trial in which certain results were achieved. This data is stored at PSNC.

## PSNC data centre

<b>Project name</b>	<b>CATALYST</b>
ID	768739
Project coordinator	Massimo Bertoncini
Project description	Converting DCs into Energy Flexibility Ecosystems
Funder	H2020-EE-2016-2017
Principal researcher	Ariel Oleksiak
Collecting period	TBD
Trial ID	TBD based on definition of use cases with the use of PSNC Pilot
Data set reference and name	<p>[TrialID/PSNC/DC]</p> <ul style="list-style-type: none"> <li>• PSNC data centre</li> <li>• High Performance Computing centre</li> </ul>
Data set description	<p>Measurement data from both facilities and IT. Data comes from 2MW data centre consisting of diverse systems – various HPC architectures, network equipment and others. Specific data depends on trials definition and monitoring capabilities of specific parts of DC.</p> <p>Type of data:</p> <ul style="list-style-type: none"> <li>• Power usage of servers and cooling;</li> <li>• Utilisation of servers;</li> <li>• Temperature of servers, room and coolant (for direct liquid cooling)</li> <li>• Possibly other parameters – to be checked.</li> </ul> <p>Source</p> <ul style="list-style-type: none"> <li>• PSNC data centre</li> </ul> <p>Volume</p> <ul style="list-style-type: none"> <li>• Size of data to be calculated depending on data required in trials</li> </ul> <p>Data and file formats</p> <ul style="list-style-type: none"> <li>• Mostly key-value pairs for various parameters</li> </ul> <p>Purpose</p> <ul style="list-style-type: none"> <li>• Analysis of real large-scale data centre operation</li> </ul>
Standards and metadata	Data is identified by a related timeframe. Other metadata to be discussed (possibly to be synchronised with a CATALYST data model).
Data sharing	Data from PSNC data centre cannot be shared outside the consortium. Currently are stored in file-based and relational databases at PSNC.
Archiving and preservation	Measurements from the PSNC data centre are stored at PSNC. This data set can provide insights for development of CATALYST framework and exploitation plans.

## SBP Pilot

SBP pilot consists of a colocation data centre integrated with offices.

Project name	CATALYST
ID	768739
Project coordinator	Massimo Bertoncini
Project description	Converting DCs into Energy Flexibility Ecosystems
Funder	H2020-EE-2016-2017
Principal researcher	Arjan Westerhoff
Collecting period	TBD
Trial ID	TBD based on definition of use cases with the use of SBP Pilot
Data set reference and name	<p>[TrialID/SBP/DC]</p> <p>Reuse the warm air to provide heat to the DC offices</p> <ul style="list-style-type: none"> <li>• Schuberg Philis office and datacentre</li> <li>• Reuse the warm air to pre-heat the diesel UPS engines</li> <li>• Provide a comparison between battery based and dynamic rotating UPS</li> </ul>
Data set description	<p>Measurement data from the data centre and office.</p> <p>Type of data:</p> <ul style="list-style-type: none"> <li>• Temperatures;</li> <li>• Cooling and heating capacity;</li> <li>• Flow</li> <li>• Electrical power specifications ( for example, kW, A, )</li> </ul> <p>Source</p> <ul style="list-style-type: none"> <li>• Building Management System and schematics</li> </ul> <p>Volume</p> <ul style="list-style-type: none"> <li>• Size of data is maximum 0.5 Gig</li> </ul> <p>Data and file formats</p> <ul style="list-style-type: none"> <li>• Mostly key-value pairs for various parameters. Data can be easily translated into the CATALYST format (partially done for monitoring in DOLFIN).</li> </ul> <p>Purpose</p> <ul style="list-style-type: none"> <li>• Research of the calculation of the correct amount of cooling and heating capacity</li> </ul>
Standards and metadata	<p>Data is identified by a related timeframe. Within CATALYST Trial ID will be used to identify data sets.</p> <p>Other metadata to be discussed (possibly to be synchronised with a CATALYST data model).</p>
Data sharing	Data from SBP datacentre can be shared.
Archiving and preservation	<p>Measurements from the data centre and office are currently stored at SBP for at least period of 1 year. The total volume depending on trials timeframe can reach up to 0.5 Gb</p> <p>This data set can demonstrate details of a trial in which certain results were achieved.</p>

## QRN Pilot

QRN pilot concentrates on a distributed DC in the form of multiple radiators deployed in residential/office buildings.

<b>Project name</b>	<b>CATALYST</b>
ID	768739
Project coordinator	Massimo Bertoncini
Project description	Converting DCs into Energy Flexibility Ecosystems
Funder	H2020-EE-2016-2017
Principal researcher	Nicolas SAINTHERANT
Collecting period	TBD
Trial ID	TBD according to CATALYST trial identifier
Data set reference and name	<p>[TrialID/Qarnot/DistributedDC]</p> <ul style="list-style-type: none"> <li>• QRN distributed cloud HPC data centre</li> <li>• Real production housing building</li> </ul>
Data set description	<p>Measurements are used for platform management and scheduling needs, according to the size of the building we can get data from 20 to 300 servers. Such installation are simply connected to local regular energy grid.</p> <p>Type of data:</p> <ul style="list-style-type: none"> <li>• Power usage of the servers</li> <li>• Temperatures (targets, dissipators, room)</li> <li>• IT monitoring: CPU utilization (%) of servers,</li> <li>• Global site metrics: Storage utilisation (%), Network utilisation (%), total power consumption of the building (if we are connected to smart power counters)</li> <li>• Energy market: prices of energy (almost static)</li> <li>• Internet bandwidth, latency, and global quality</li> <li>• Cap on the total energy we are allowed to use, or the building is allowed to use (almost static)</li> </ul> <p>Source</p> <ul style="list-style-type: none"> <li>• One housing building (TBD)</li> </ul> <p>Volume</p> <ul style="list-style-type: none"> <li>• TBD Depends on the time frequency...</li> </ul> <p>Data and file formats</p> <ul style="list-style-type: none"> <li>• TBD</li> </ul>
Standards and metadata	<p>Data is identified by a related timeframe. Within CATALYST Trial ID will be used to identify data sets.</p> <p>Other metadata to be discussed (possibly to be synchronised with a CATALYST data model).</p>
Data sharing	<p>Data from QRN distributed DC cannot be shared as-is, primarily for user privacy reasons. It enclose personal data and as such should be treated cautiously, even if these data cannot be tied to one person or household.</p> <p>The data sets will be shared only within the CATALYST project members upon request.</p> <p>About data protection, this data sets will be accessible to authorized person at Qarnot, mainly researchers working on the CATALYST project.</p>
Archiving and preservation	<p>Similar data sets are stored in data centre, the CATALYST trial data set will most probably be stored in a regular data centre as well. Again, the end volume will depend on many parameters, but especially the frequency and the time frame.</p>

	The storage costs will be covered by Qarnot as regular data sets storage. The data sets will be kept during all the project duration and during 3 years after at least. Depending on further use of this data, data protection and storage policy may evolve after the project.
--	---

## 6.2.2 Social data sets

### GDC-SG data set

Data set related to the GDC-SG.

Project name	CATALYST
ID	768739
Project coordinator	Massimo Bertoncini
Project description	Converting DCs into Energy Flexibility Ecosystems
Funder	H2020-EE-2016-2017
Principal researcher	Vasiliki Georgiadou (GIT)
Collecting period	During the project's lifetime
Trial ID	N/A
Data set reference and name	<ul style="list-style-type: none"> <li>Information on members of the Green Data Centre – Stakeholder Group (GDC-SG) and the project's Advisory Board. The latter is essentially a subset of the former.</li> <li>This data set is related to Task 8.5</li> </ul>
Data set description	<p>Type of data, e.g.</p> <ul style="list-style-type: none"> <li>Personal info such as, photo, name, job description and affiliation, contact info (such as email), possible social media profile info on LinkedIn and Twitter.</li> <li>Their activity and participation to GDC-SG meetings, events and related tasks. This includes for example meeting minutes, event proceedings, photos, presentations and conference calls records.</li> </ul> <p>Source</p> <ul style="list-style-type: none"> <li>The members interested in the GDC-SG</li> </ul> <p>Volume</p> <ul style="list-style-type: none"> <li>Depends on the number of members, related events and meetings.</li> </ul> <p>Data and file formats</p> <ul style="list-style-type: none"> <li>As applicable, for example an excel file for the members directory, meeting minutes in doc or pdf format.</li> </ul>
Standards and metadata	TBD at a later stage once we obtain more information and experience on the GDC-SG membership and operations
Data sharing	<p>Which data will be shared?</p> <ul style="list-style-type: none"> <li>The directory of GDC-SG members will be public at all times</li> <li>Private meeting minutes will be shared only within the GDC-SG</li> </ul>

	<ul style="list-style-type: none"> <li>Public events proceedings, presentations, videos and such will be public</li> </ul> <p>Data protection</p> <ul style="list-style-type: none"> <li>The whole consortium will have access to these data at all times.</li> </ul>
Archiving and preservation	After the completion of the project, the GDC-SG budget should allocate funding for this.



## 7 Conclusions

This document provides the first approach to the CATALYST data management plan. It delivers a template for data descriptions with initial definitions of possible data sets. It also contains the first proposal of data management methodology along with data protection measures.

In the next version of the DMP, data sets will be defined in more detail based on definitions of trials, which should be ready by M18. Final data formats will also be proposed as data models and monitoring software should be completed by that time.



## 8 References

- [1] European Commission, “Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, Version 1.0,” December 2013. [Online]. Available: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hioa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hioa-pilot-guide_en.pdf).
- [2] European Commission, “Guidelines on Data Management in Horizon 2020, Version 1.0,” December 2013. [Online]. Available: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hioa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hioa-data-mgt_en.pdf).
- [3] “OpenAIRE,” [Online]. Available: <https://www.openaire.eu/open-access-in-horizon-2021>.
- [4] V. Avelar, D. Azevedo, A. French and E. N. Power, “Pue: A comprehensive examination of the metric,” *The Green Grid*, 2012, white paper 49.
- [5] CENELEC - EN 50600-1, “INFORMATION TECHNOLOGY - DATA CENTRE FACILITIES AND INFRASTRUCTURES - PART 1: GENERAL CONCEPTS,” [Online]. Available: <https://standards.globalspec.com/standards/detail?familyId=NKXDSEAAAAAAAAA>.
- [6] ASHRAE technical committee (TC) 9.9, “Thermal guidelines for data processing environments expanded data center classes and usage guidance,” vol. 9, 2011.
- [7] H2020 CATALYST project, “Deliverable D8.3 - Report on Marketing & Promotional Tools V1,” 2018.
- [8] H. C. project, “Deliverable D8.1 - Report to the project’s website”.
- [9] T. C. I. A. C. P. a. I. S. Marcel Antal, “Transforming Data Centers in Active Thermal Energy Players in Nearby Neighborhoods,” *Sustainability*, vol. 10, no. 4, 2018.
- [10] I. A. I. S. M. A. C. P. M. B. D. A. F. P. Tudor Cioara, “Exploiting data centres energy flexibility in smart cities: Business scenarios,” *Information Sciences*, no. ISSN 0020-0255, 2018.
- [11] T. V. M. B. A. V. A. O. N. S. V. G. I. A. M. A. P. C. P. Tudor Cioara, “Converting Data Centres in Energy Flexibility Ecosystems,” in *IEEE 18th International Conference on Environment and Electrical Engineering and 2nd Industrial and Commercial Power Systems Europe*, Palermo, 2018.
- [12] E. P. M. A. C. P. C. T. I. A. a. I. S. Cristian Pinteau, “CoolCloudSim: Integrating Cooling System Models in CloudSim,” in *ICCP*, Cluj-Napoca, 2018.
- [13] [Online]. Available: [http://project-catalyst.eu/wp-content/uploads/2018/06/GDC-SG\\_Brochure\\_Web.pdf](http://project-catalyst.eu/wp-content/uploads/2018/06/GDC-SG_Brochure_Web.pdf).

- [14] [Online]. Available: <https://datacenterworks.nl/2018/08/10/een-datacenter-is-veel-meer-dan-enkel-een-plek-om-data-te-verwerken/>.
- [15] [Online]. Available: <https://www.eventbrite.com/e/green-data-centre-stakeholders-group-kick-off-tickets-45548811722>.
- [16] [Online]. Available: <https://www.innogrid2020.eu/>.
- [17] [Online]. Available: <https://www.publishing.ninja/V4/page/5499/279/19/1#>.
- [18] [Online]. Available: <http://www.greenitamsterdam.nl/news/1227-introducing-the-catalyst-to-convert-data-centres-into-energy-flexibility-ecosystems> .
- [19] [Online]. Available: <https://thedatacenter.blog/catalyst/> .
- [20] [Online]. Available: <https://www.greendatacenterplatform.com/project/green-it-amsterdam-catalyst-2/> .
- [21] [Online]. Available: <https://twitter.com/Qarnot>.
- [22] [Online]. Available: <https://blog.qarnot.com/convertng-data-centres-in-energy-flexibility-ecosystems/> .
- [23] [Online]. Available: <https://dca-global.org/news/view/2563/the-green-data-centre-stakeholders-group-kicks-off> .
- [24] [Online]. Available: <https://amsterdamsmartcity.com/projects/catalyst> .
- [25] C. Consortium, “Deliverable: D2.2 - Green Data Centre Assessment Toolkit,” 2018.
- [26] “European task force for the implementation of smart grids into the European internal market,” [Online]. Available: [https://ec.europa.eu/energy/sites/ener/files/documents/mission\\_and\\_workprogramme.pdf](https://ec.europa.eu/energy/sites/ener/files/documents/mission_and_workprogramme.pdf).
- [27] “Smart grids task force (SGTF) workshop on experiences and conditions for successful implementation of smart homes and buildings,” [Online]. Available: [https://ec.europa.eu/energy/sites/ener/files/documents/Agenda%20table%20SGTF-EG3%20%20Smart%20Homes\\_%2026%2](https://ec.europa.eu/energy/sites/ener/files/documents/Agenda%20table%20SGTF-EG3%20%20Smart%20Homes_%2026%2).
- [28] “Smart Grid Task Force (SGTF),” [Online]. Available: <https://www.edsoforsmartgrids.eu/policy/eu-steering-initiatives/smart-grid-task-force/>.
- [29] “Smart Grid projects in Europe: lessons learned and current developments,” [Online]. Available: [https://s3.amazonaws.com/academia.edu.documents/8342518/jrc\\_report\\_-\\_smart\\_grid\\_projects\\_in\\_europe.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1536825011&Signature=pJ6pACUljQTOI](https://s3.amazonaws.com/academia.edu.documents/8342518/jrc_report_-_smart_grid_projects_in_europe.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1536825011&Signature=pJ6pACUljQTOI).

- [30] “OpenADR,” [Online]. Available: <https://www.openadr.org/about-us> .
- [31] “The OpenADR Primer,” [Online]. Available: [https://www.openadr.org/assets/docs/openadr\\_primer.pdf](https://www.openadr.org/assets/docs/openadr_primer.pdf).
- [32] “Open Automated Demand Response Communications Specification (Version 1.0)”.
- [33] [Online]. Available: <http://www.auto-grid.com/>.
- [34] “OpenADR Certified Products,” [Online]. Available: <https://products.openadr.org/> .
- [35] “Smart Grids European Technology Platform,” [Online]. Available: [https://www.earpa.eu/earpa/39/etp\\_smartgrids.html](https://www.earpa.eu/earpa/39/etp_smartgrids.html) .
- [36] “European Smart Grids Technology Platform Vision and Strategy,” [Online]. Available: [https://ec.europa.eu/research/energy/pdf/smartgrids\\_en.pdf](https://ec.europa.eu/research/energy/pdf/smartgrids_en.pdf) .
- [37] “Smart Grid European Technology Platform working groups,” [Online]. Available: <https://www.etip-snet.eu/about/working-groups/> .
- [38] “Intensys4EU Project,” [Online]. Available: <https://www.etip-snet.eu/intensys4eu/> .
- [39] “European Technology and Innovation Platform on Deep Geothermal,” [Online]. Available: <https://www.etip-dg.eu/> .
- [40] “European Technology and Innovation Platform for Ocean Energy,” [Online]. Available: <https://www.etipocean.eu/> .
- [41] “ETP Smart Grids Vision for 2050,” [Online]. Available: <https://www.etip-snet.eu/wp-content/uploads/2018/05/ETIP-SNET-Presentation-27-June.pdf>.
- [42] “ETP 2017-2020,” [Online]. Available: <https://www.etip-snet.eu/wp-content/uploads/2017/10/ETIP-SNET-Implementation-Plan-2017-2020.pdf> .
- [43] “Smart Energy Demand Coalition Position Paper,” [Online]. Available: <http://www.smartenergy.eu/wp-content/uploads/2011/03/Smart-Energy-Demand-Coalition-Position-Paper-Final.pdf> .
- [44] “Position Paper: Smart Metering and Information Smart Meters and their central role in consumer empowerment,” [Online]. Available: <http://www.smartenergy.eu/wp-content/uploads/2012/02/SEDC-Smart-Meter-Position-Paper.pdf> .
- [45] “Data Privacy Position Paper: ENERGY USAGE DATA ACCESS: A POLICY THAT PROTECTS CONSUMERS WHILE DELIVERING MAXIMUM BENEFITS,” [Online]. Available: <http://www.smartenergy.eu/wp-content/uploads/2011/06/SEDC-Data-Privacy-Position-Paper.pdf>.

- [46] “Monitoring Demand-Side Flexibility Assessing the potential and market uptake of Demand-Side Flexibility in the European energy system,” [Online]. Available: <http://www.smart-en.eu/wp-content/uploads/2016/07/SEDC-Monitoring-Dem>.
- [47] “Demand Response at the DSO level Enabling DSOs to harness the benefits of demand-side flexibility,” [Online]. Available: <http://www.smart-en.eu/wp-content/uploads/2016/05/SEDC-White-Paper-Demand-Response-at-the-DSO-level.pdf> .
- [48] “10 Recommendations for an Efficient European Power Market Design,” [Online]. Available: <http://www.smart-en.eu/wp-content/uploads/2016/02/SEDC-10-recommendations.pdf> .
- [49] “A new market design for competition, innovation and consumer empowerment,” [Online]. Available: <http://www.smart-en.eu/wp-content/uploads/2018/05/SEDC-Market-Design-position-paper-6-June-2017.pdf> .
- [50] “Policy Recommendations for trilogue negotiations on the Electricity Market Design,” [Online]. Available: <http://www.smart-en.eu/wp-content/uploads/2018/05/smartEn-Policy-recommendations-EMD-final.pdf> .
- [51] “Work Programme 2016 – CEN-CENELEC,” [Online]. Available: [https://www.cenelec.eu/news/publications/Publications/CEN-CENELEC-WP2016\\_EN.pdf](https://www.cenelec.eu/news/publications/Publications/CEN-CENELEC-WP2016_EN.pdf) .
- [52] “CEN – Technical Bodies – CEN/CLC/ETSI/SMCG,” [Online]. Available: [https://standards.cen.eu/dyn/www/f?p=204:7:0:::FSP\\_ORG\\_ID:881455&cs=1BE8DC5785D0489441230DFFAAF473980](https://standards.cen.eu/dyn/www/f?p=204:7:0:::FSP_ORG_ID:881455&cs=1BE8DC5785D0489441230DFFAAF473980).
- [53] “Standardisation landscape for the energy management and environmental viability of data centres,” [Online]. Available: [ftp://ftp.cenelec.eu/EN/EuropeanStandardization/HotTopics/ICT/GreenDataCentres/GDC\\_landscape\\_Ed4\(2017\).pdf](ftp://ftp.cenelec.eu/EN/EuropeanStandardization/HotTopics/ICT/GreenDataCentres/GDC_landscape_Ed4(2017).pdf) .
- [54] “Energy Management and Environmental Viability of Data Centres,” [Online]. Available: [ftp://ftp.cenelec.eu/EN/EuropeanStandardization/HotTopics/ICT/GreenDataCentres/GDC\\_Brochure\\_Ed4\(2017\).pdf](ftp://ftp.cenelec.eu/EN/EuropeanStandardization/HotTopics/ICT/GreenDataCentres/GDC_Brochure_Ed4(2017).pdf) .
- [55] “Welcome to the IEC – International Electrotechnical Commission,” [Online]. Available: <http://www.iec.ch/> .
- [56] “IEC – SyC Smart Energy Dashboard > Projects: Work programme, Publications,” [Online]. Available: [http://www.iec.ch/dyn/www/f?p=103:38:5861932394203:::FSP\\_ORG\\_ID,FSP\\_APEX\\_PAGE,FSP\\_PROJECT\\_ID:11825,23,101948](http://www.iec.ch/dyn/www/f?p=103:38:5861932394203:::FSP_ORG_ID,FSP_APEX_PAGE,FSP_PROJECT_ID:11825,23,101948).
- [57] “SC C5 Electricity Markets and Regulation,” [Online]. Available: <http://c5.cigre.org/>.
- [58] *JWG C2/C5-5 - Survey of Power Exchanges - Organization and Responsibilities, 2007 CIGRE Regional Meeting, Tallin, Estonia.*

- [59] JWG C2/C5-5 - *Interfaces Between System Operators, Market Operators and other Market Actors*, 2006, Paris General Meeting.
- [60] JWG C2/C5-5 - *Interfaces Between System Operators, Market Operators and other Market Actors*, 2006, Paris General Meeting.
- [61] “The Alliance for Internet of Things Innovation (AIOTI),” [Online]. Available: <https://www.iotone.com/organization/the-alliance-for-internet-of-things-innovation-aioti/o218> .
- [62] “IETF Energy Management Working Group,” [Online]. Available: <https://datatracker.ietf.org/wg/eman/about/>.
- [63] “EMAN: Energy-Management Activities at the IETF report,” [Online]. Available: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6509877>.
- [64] “EMAN published standards,” [Online]. Available: <https://datatracker.ietf.org/wg/eman/documents/>
- [65] “DOLFIN Whitepaper,” [Online]. Available: <http://www.dolphin-fp7.eu/wp-content/uploads/2014/01/WHITEPAPER-DOLFIN-final.pdf>.
- [66] “WELMEC Information,” [Online]. Available: <https://www.welmec.org/welmec/welmec-tour/welmec-information/>.
- [67] “WELMEC Strategy,” [Online]. Available: [https://www.welmec.org/fileadmin/user\\_files/pdf/WELMEC\\_strategy\\_2015-2020.pdf](https://www.welmec.org/fileadmin/user_files/pdf/WELMEC_strategy_2015-2020.pdf).
- [68] “WELMEC - An introduction,” [Online]. Available: [https://www.welmec.org/fileadmin/user\\_files/publications/WELMEC-general/WELMEC\\_Guide\\_1-2017\\_-\\_WELMEC\\_An\\_introduction.pdf](https://www.welmec.org/fileadmin/user_files/publications/WELMEC-general/WELMEC_Guide_1-2017_-_WELMEC_An_introduction.pdf) .
- [69] “WELMEC Software Guide,” [Online]. Available: [https://www.welmec.org/fileadmin/user\\_files/publications/WG\\_07/Guide\\_7.2\\_2015\\_\\_Software.pdf](https://www.welmec.org/fileadmin/user_files/publications/WG_07/Guide_7.2_2015__Software.pdf).
- [70] “USEF Foundation,” [Online]. Available: <https://www.usef.energy/usef-foundation/>.
- [71] “Arnoud Backers et al. An introduction to the Universal Smart Energy Framework”.
- [72] “USEF Publications,” [Online]. Available: <https://www.usef.energy/news-events/publications/>.
- [73] “EF,” [Online]. Available: <https://flexible-energy.eu/efi/>.
- [74] “EFI 2.0 Specification,” [Online]. Available: <https://github.com/flexiblepower/efi/blob/master/specification/EFI%202.0%20Specification.pdf> .

- [75] [Online]. Available: <https://dca-global.org/news/view/2563/the-green-data-centre-stakeholders-group-kicks-off>.
- [76] [Online]. Available: <https://datacenterworks.nl/2018/07/03/green-datacenter-stakeholder-group-van-start/>.
- [77] [Online]. Available: <https://www.computerweekly.com/news/252446170/EU-backed-datacentre-consortium-seeks-industry-help-for-Europe-wide-sustainability-push>.
- [78] [Online]. Available: [https://www.economie.gouv.qc.ca/objectifs/informer/vecteurs/vecteurs-actualites/vecteurs-actualites-details/?no\\_cache=1&tx\\_ttnews%5Btt\\_news%5D=22841&tx\\_ttnews%5Bcat%5D&cHash=4a5f219472ba509de567d91176525526](https://www.economie.gouv.qc.ca/objectifs/informer/vecteurs/vecteurs-actualites/vecteurs-actualites-details/?no_cache=1&tx_ttnews%5Btt_news%5D=22841&tx_ttnews%5Bcat%5D&cHash=4a5f219472ba509de567d91176525526).
- [79] [Online]. Available: <https://datacenterworks.nl/2018/07/03/green-datacenter-stakeholder-group-van-start/>.
- [80] [Online]. Available: <https://www.computerweekly.com/news/252446170/EU-backed-datacentre-consortium-seeks-industry-help-for-Europe-wide-sustainability-push>.
- [81] [Online]. Available: [https://www.economie.gouv.qc.ca/objectifs/informer/vecteurs/vecteurs-actualites/vecteurs-actualites-details/?no\\_cache=1&tx\\_ttnews%5Btt\\_news%5D=22841&tx\\_ttnews%5Bcat%5D&cHash=4a5f219472ba509de567d91176525526](https://www.economie.gouv.qc.ca/objectifs/informer/vecteurs/vecteurs-actualites/vecteurs-actualites-details/?no_cache=1&tx_ttnews%5Btt_news%5D=22841&tx_ttnews%5Bcat%5D&cHash=4a5f219472ba509de567d91176525526).

