



Project Acronym:	VICINITY
Project Full Title:	Open virtual neighbourhood network to connect intelligent buildings and smart objects
Grant Agreement:	688467
Project Duration:	48 months (01/01/2016 - 31/12/2019)

Deliverable D9.3

Data Management Plan, second version

Work Package:	WP9 – Dissemination of Results & Exploitation
Task(s):	T9.2 – Data Management Plan
Lead Beneficiary:	HITS
Due Date:	31 December 2017 (M24)
Submission Date:	22 December 2017 (M24)
Deliverable Status:	Final
Deliverable Type:	R
Dissemination Level:	PU
File Name:	VICINITY_D9.3_Data_Management_Plan_v1_0.pdf

VICINITY Consortium

No	Beneficiary		Country
1.	TU Kaiserslautern (Coordinator)	UNIKL	Germany
2.	ATOS SPAIN SA	ATOS	Spain
3.	Centre for Research and Technology Hellas	CERTH	Greece
4.	Aalborg University	AAU	Denmark
5.	GORENJE GOSPODINJSKI APARATI D.D.	GRN	Slovenia
6.	Hellenic Telecommunications Organization S.A.	OTE	Greece
7.	bAvenir s.r.o.	BVR	Slovakia
8.	Climate Associates Ltd	CAL	United Kingdom
9.	InterSoft A.S.	IS	Slovakia
10.	Universidad Politécnica de Madrid	UPM	Spain
11.	Gnomon Informatics S.A.	GNOMON	Greece
12.	Tiny Mesh AS	TINYM	Norway
13.	HAFENSTROM AS	ITS	Norway
14.	Enercoutim – Associação Empresarial de Energia Solar de Alcoutim	ENERC	Portugal
15.	Municipality of Pylaia-Hortiatis	MPH	Greece

Authors List

Leading Author (Editor)				
Surname	First Name	Beneficiary	Contact email	
Sveen	Flemming	HITS	flsveen@online.no	
Co-authors (in alphabetic order)				
No	Surname	First Name	Beneficiary	Contact email
1.	Vásquez	Juan C.	AAU	juq@et.aau.dk
2.	Hovstø	Asbjørn	HITS	hovsto@online.no
3.	Kaggelides	Konstantinos	GNOMON	k.kaggelides@gnomon.com.gr
4.	Samovich	Natalie	ENERC	n.samovich@enercoutim.eu
5.	Tryferidis	Athanasios	CERTH	thanasic@iti.gr
6.	Zandes	Dimitris	GNOMON	d.zandes@gnomon.com.gr

Reviewers List

List of Reviewers (in alphabetic order)				
No	Surname	First Name	Beneficiary	Contact email
1.	Chochliouros	Ioannis	OTE	ichochliouros@otereseach.gr
2.	Zandes	Dimitris	GNOMON	d.zandes@gnomon.com.gr
3.	Heinz	Christopher	UNIKL	heinz@cs.uni-kl.de

Revision Control

This document is based on D9.2 Data Management Plan (year 1). D9.3 contains some major upgrades to the datasets as the planning has matured. There have also been made some changes to information about the pilot sites. A new section addresses security issues and open calls. This is currently tentative, but will currently serves as a guiding tool. SysML – a modelling language is described, and there have also been made some updates on ethics and security. Finally, the ethics advisory board has been a new addition to this document.

Version	Date	Status	Modifications made by
0.1	4. September 2017	Initial Draft	Sveen (HITS)
0.2	1. November 2017	First Draft formatted with contributions received	Tryferidis (CERTH), Juan (AAU), Sveen (HITS)
0.3	10. November 2017	Deliverable version for final review by partners	Tryferidis (CERTH), Juan (AAU), Olav (TINYM), Sveen (HITS), DIMITRIS (GNOMON)
0.4	24. November 2017	Final improvements	Sveen (HITS), Saleiro (ENERC), Oliveira (ENERC)
0.5	1. December 2017	Deliverable version uploaded for Quality Check	Sveen (HITS)
0.5.1	8. December 2017	Deliverable version uploaded for Quality Check (2 nd version)	Sveen (HITS), Hovstø (HITS)
0.6	15. December 2017	Quality Check	Sveen (HITS), Zandes (Gnomon)
0.7	22. December 2017	Final improvements	Sveen (HITS)
1.0	22. December 2017	Submission to the EC	Sveen (HITS)

List of Definitions and Abbreviations (A-R)

Abbreviation	Definition	Abbreviation	Definition
AI	Artificial Intelligence	HTTP	Hypertext Transport Protocol
API	Application Programming Interface	ICT	Information & Communication Technologies
ASN	Abstract Syntax Notation	IEQ	Indoor Environmental Quality
ATDD	Acceptance Test Driven Development	IoT	Internet of Things
BDD	Behavior Driven Development	ICT	Information & Communication Technologies
CA	Consortium Agreement	IP	Internet Protocol
CL	Classified	IPR	Intellectual Property Right
CO	Confidential	IR	Infrared
CoAP	Constrained Application Protocol	ITU	International Telecommunication Union
DER	Distributed Energy Resources	M2M	Machine to Machine
DG	Distribution Grid	MQTT	Message Queuing Telemetry Transport
DL	Description logic	NFC	Near Field Communication
DSM	Digital Single Market	OSG	Open Geospatial Consortium
DSM	Demand Side Management	OS	Operating System
DSO	Distribution System Operator	OWL	Web Ontology Language
EC	European Commission	PBT	Property Based Testing
ESCO	Energy Service Company	PC	Project Coordinator
ESO	European Standards Organization	PO	Project Officer
ETSI	European Telecommunications Standards Institute	PU	Public
EU	European Union	QA	Quality Assurance
GDPR	General Data Protection Regulation	RDF	Resource Description Framework
GPS	Geographic Positioning System	RES	Renewable Energy Resources
GSM	Global System for Mobile communications	RFID	Radio Frequency Identification

List of Definitions and Abbreviations (S-X)

Abbreviation	Definition	Abbreviation	Definition
SG	Study Group	TDD	Test Driven Development
SM	Scientific Manager	TSO	Transmission System Operator
SME	Small Medium Enterprise	UA	Unified Architecture
SOA	Service Oriented Architecture	UDP	User Datagram Protocol
SSL	Secure Socket Layer	URL	Uniform Resource Locator
SSN	Social Security Number / Semantic Sensor Network	UUID	Universally unique identifier
SW	Software	VNM	Vicinity Neighbourhood Manager
TCP	Transmission Control Protocol	WP	Work Package
		XML	EXtensible Markup Language

Content

VICINITY Consortium	2
Authors List	3
Reviewers List	3
Revision Control	4
List of Definitions and Abbreviations (A-R)	5
List of Definitions and Abbreviations (S-X)	6
1. Executive Summary	9
2. Introduction	10
3. General Principles	12
3.1. Participation in the Pilot on Open Research Data	12
3.2. IPR management and security	12
3.3. Personal Data Protection	12
3.4. Production data	15
3.5. Ethics and security	16
3.6. The VICINITY Data Management Portal	17
3.7. Format of datasets	18
DS. PARTiCiPANTName.##.Logical_sensorname	18
3.8. Open Call	19
3.9. Description of methods for dataset description	20
3.10. Standards and metadata	21
3.11. Data sharing	22
3.12. Archiving and preservation (including storage and backup)	23
4. Datasets for smart grid from Aalborg University (AAU)	24
DS.AAU.01.GRID_Status	24
5. Datasets for smart energy from Enercutim (ENERC)	26
DS.ENERC.01.METEO_Station	26
DS.ENERC.02.BUILDING_Status.....	27
DS.ENERC.03.GRID_Status.....	29
6. Datasets for eHealth from GNOMON Informatics SA (GNOMON)	31
DS.GNOMON.01.Pressure_sensor.....	31
DS.GNOMON.02.Weight_sensor	33
DS.GNOMON.03.Fall_sensor	34
DS.GNOMON.04.Wearable_Fitness_Tracker_Sensor	36
DS.GNOMON.05.Beacon_Sensor	38
DS.GNOMON_CERTH.06.Gorenje_Smart_Appliances_Sensor	40
7. Datasets for eHealth from Centre for Research and Technology Hellas (CERTH)	42
DS.CERTH.01.Occupancy_Sensor	42
DS.CERTH.02.Motion_Sensor	43
8. Datasets for intelligent mobility from Hafenstrom AS (HITS)	46
DS.HITS.01.Parkingsensor	46
DS.HITS.02.SmartLight.....	47

DS.HITS.03.LaptopTeststation48

DS.HITS.04.Sensio_sensors_temperature_motion_lock50

DS.HITS.05.Gorenje_Smart_Appliances_Sensor51

9. Datasets for buildings from Tiny Mesh AS (TINYM)..... 53

DS. TinyMesh.01.Door_Sensor53

DS. TinyMesh.02.Energy_Water_Consumption_Sensor54

DS. TinyMesh.03 Tinymesh_Gateway55

10. Conclusions 57

References 58

Annex 1 – Preferred Formats 59

3.13. Selection of File Formats59

3.14. Preferred and Acceptable Formats59

Annex 2: VICINITY Consent form Template 61

Annex 3 – The Ethical Advisory Board 62

Annex 4 – Contacts technical data and internal training 63

Annex 5 – Assessment tools 64

List of Tables 66

List of Figures 66

1. Executive Summary

«The VICINITY project will build and demonstrate a bottom-up ecosystem of decentralised interoperability of IoT infrastructures called virtual neighborhood, where users can share the access to their smart objects without losing the control over them.»

The present document is a deliverable “D9.3 – Data Management Plan” of the VICINITY project (Grant Agreement No.: 688467), funded by the European Commission’s Directorate-General for Research and Innovation (DG RTD), under its Horizon 2020 Research and Innovation Programme (H2020).

The VICINITY Consortium has identified several areas that need to be addressed; Protocol interoperability, identification tokens, encryption keys, data formats and packet size. Also, several issues are related to latency, bandwidth and general architecture.

VICINITYs activities will involve human participants, as some of the pilots will be conducted in real homes with actual residents. For some of the activities to be carried out by the project, it may be necessary to collect basic personal data (e.g. name, background, contact details), even though the project will avoid collecting such data unless necessary. Such data will be protected in accordance with the EU’s Data Protection Directive 95/46/EC¹ of the European Parliament and of the Council of 24th of October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. National and local legislations applicable to the project will also be strictly applied (full list described in annex 2: ethics and security).

All personal data, or data directly related to the residents, will first be collected when the project has received a signed informed consent form from the subjects participating.

This is the second version of the project Data Management Plan (DMP). It contains preliminary information about the 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 purpose of the Data Management Plan (is to provide an analysis of the main elements of the data management policy that will be used by the consortium with regard to all the datasets that will be generated by the project. The DMP is not a fixed document, but will evolve during the lifespan of the project (Figure 1).

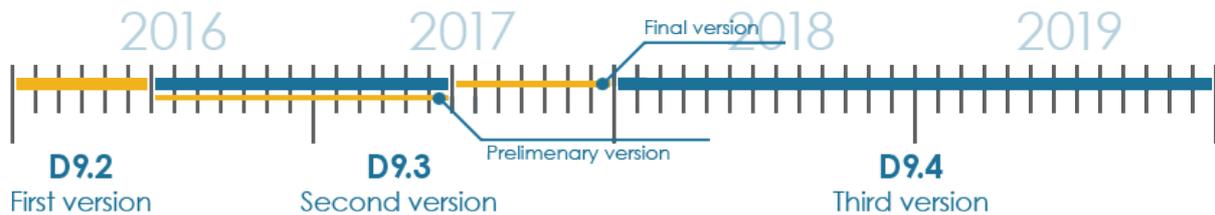


Figure 1: Data Management Plan – deliverables 2016 – 2019

Note: In order to assist the official project review process by the commission for the first project period (M1-M24), a preliminary version of the updated DMP of D9.3 was delivered prior to M24 (December 2017), in order to be enable a better assessment of the progress of the Data Management in the project by the reviewers.

¹ <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en>

The datasets referred to in this document are drafted during the first project stages (completed 30th of June 2016) of the project. The document can only reflect the intentions of the project partners toward developing the overall project's datasets. The second revision (D9.3) has been prepared for 31st December 2017, and the third (D9.4) will be ready by 31st December 2019. This follows the H2020 guidelines on Data Management Plans, and as stated in the Grant Agreement 688467.

As the project progresses and results start to arrive, the datasets will be elaborated on. The detailed descriptions of all the specific datasets that have been collected will be described, made available under the relevant Data Management framework.

2. Introduction

The purpose of the Data Management Plan (DMP) deliverable is to provide relevant information concerning the data that will be collected and used by the partners of the project VICINITY. The project aims to develop a solution defined as "Interoperability as a Service" which will be a part of the VICINITY open gateway (Figure 2). In order to achieve this, a platform for harvesting, converting and sharing data from IoT units has to be implemented on the service layer of the network.

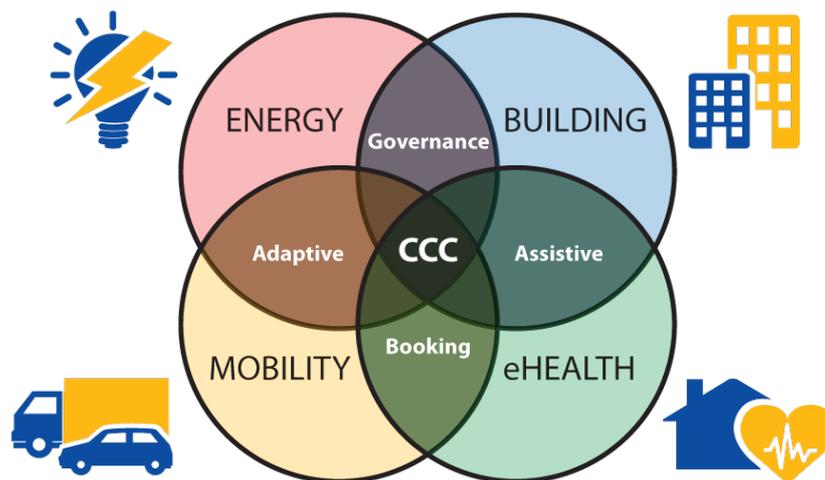


Figure 2: Domains and some of the functionalities the DMP has to cover

This goal entails the need for good documentation and implementation of descriptors, lookup-tables, privacy settings and intelligent conversion of data formats. The strength of having a cloud-based gateway is that it should be relatively simple to upgrade with new specifications and implement conversion, distribution and privacy strategies. In particular, the privacy part is considered an important aspect of the project, as VICINITY needs to follow and adhere to strict privacy policies. It will also be necessary to focus on possible ethical issues and access restrictions regarding personal data so that no regulations on sensitive information are violated.

The datasets collected will belong to four main domains; smart energy, mobility, smart home and eHealth (Figure 3: Example of potential data points in use cases that generate data.). There exist several standards and guidelines the project needs to be aware within each of these fields. There are a number of different vendors and disciplines involved – and much of the information that is available only exists in proprietary data formats. For this reason, VICINITY will target IoT units that follow the specifications defined by oneM2M consortium, ETSI standardization group and international groups and committees.

The DMP has been undergone some changes in particular in regards to privacy concerns when collecting and distributing. This version of the document is based on the knowledge generated through discussions, demonstrations and preparations for deployment at pilot sites.

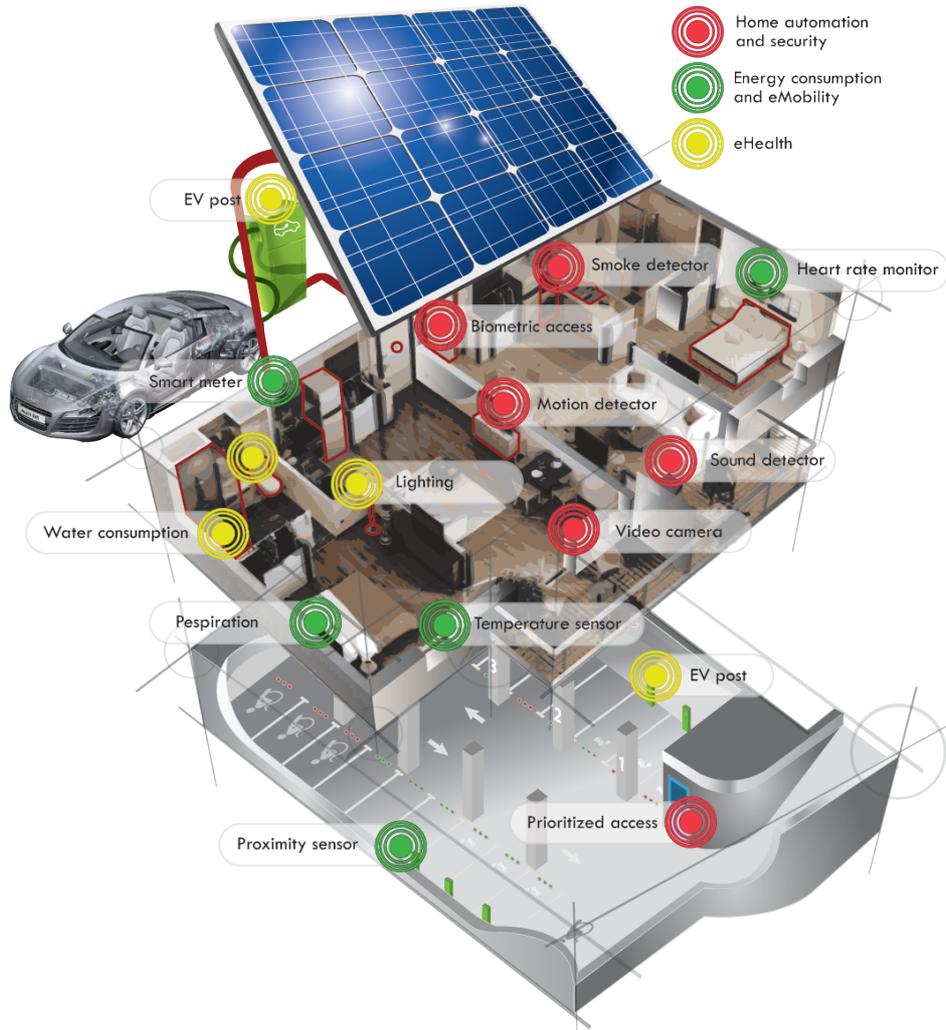


Figure 3: Example of potential data points in use cases that generate data.

3. General Principles

3.1. Participation in the Pilot on Open Research Data

VICINITY participates in the Pilot on Open Research Data launched by the European Commission along with the Horizon2020 programme. The consortium believes firmly in the concepts of open science, and the large potential benefits the European innovation and economy can draw from allowing reusing data at a larger scale. Therefore, all data produced by the project may be published with open access – though this objective will obviously need to be balanced with the other principles described below.

3.2. IPR management and security

As a research and innovation action, VICINITY aims at developing an open framework and gateway – but with support for value added services and business models. The project consortium includes partners from private sector, public sector and end-users (Figure 4). Some partners may have Intellectual Property Rights on their technologies and data. Consequently, the VICINITY consortium will protect that data and crosscheck with the concerned partners before data publication.



Figure 4: The VICINITY consortium includes partners from different sectors with confidential data

A holistic security approach will be followed, in order to protect the pillars of information security (confidentiality, integrity, availability). The security approach will consist of a methodical assessment of security risks followed by their impact analysis. This analysis will be performed on the personal information and data processed by the proposed system, their flows and any risk associated to their processing.

Security measures will include secure protocols (HTTPS and SSL), login procedures, as well as protection against bots and other malicious attacks such as CAPTCHA technologies. Moreover, the industrial demo sites apply monitored and controlled procedures related to the data collection, their integrity and protection. The data protection and privacy of personal information will include protective measures against infiltration as well as physical protection of core parts of the systems and access control measures.

3.3. Personal Data Protection

The technical implementation of VICINITY does not expose, use or analyze data, but some activities will involve human participants. The pilots will be conducted in real apartments and cover real use scenarios related to health monitoring, booking, home management, governance, energy consumption and other various human activity and behavior analysis –related data gathering purposes. Some of the activities to be carried out by the project may need to gather some basic personal data (e.g. name, background, contact details, interest, IoT units and assigned actions), even though the project will avoid collecting such data unless data is really necessary for the application.

Such data will be protected in accordance with the EU's Data Protection Directive 95/46/EC² “on the protection of individuals with regard to the processing of personal data and on the free movement of such data” and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (Figure 5).



Figure 5: VICINITY complies with European and national legislations

WP3 and WP4 activities dealing with the implementation and deployment of core components will be performed in Slovakia under leadership of local partners (BVR and IS). For this reason the solution will be reviewed for compliance with Data Protection Act No. 122/2013 approved by National Council of the Slovak Republic together with its amendment No. 84/2014 which already reflects the EC directive proposal 2012/0011/COD.

WP7 and WP8 activities will be performed in Greece, Portugal and Norway under the leadership of local partners. In the following the consortium outlines the legislation for the countries involved in the Trial:

- I. Greek Trial in Municipality of Pilea-Hortiatis, Thessaloniki, for Greece, legislation includes “Law 2472/1997 (and its amendment by Law 3471/2006) of the Hellenic Parliament”.
 - Regulatory authorities and ethical committees: Hellenic Data Protection Authority <http://www.dpa.gr/>
- II. Norwegian trials in Teaterkvarteret healthcare assisted living home in Tromsø and offices in Oslo Sciencepark, Oslo, have to comply with national legislation “Personal Data Act of 14 April No.31”⁵relating to the processing of personal data.
 - Each pilot demonstration has to notify regulatory body Datatilsynet pursuant to section 31 of the Personal Data Act and section 29 of the Personal Health Data Filing System Act.
- III. Portuguese Trial in Martim Longo microgrid pilot site in the Algarve region, Portugal. The Portuguese renewable energy legislative base dates back to 1988, and was upgraded and reviewed multiple times since then. The most important legislative diplomas are listed; DL 189/88, DL 168/99, DL 312/2001, DL 68/2002, DL 29/2006 and DL 153/2014. The last on the list refers to also one of the most important legislative changes, being the legislative base for broad based auto-consumption, with possibility to inject excess energy in to the grid under certain conditions.
 - The collection and use of personal data in Portugal are regulated by the following two laws: “Law 41/2004” (and its amendment “Law 46/2012”), and “Law 32/2008”.

Further information on how personal data collection and handling should be approached in the VICINITY project will be provided in other deliverables.

All personal data collection efforts of the project partners will be established after giving subjects full details on the experiments to be conducted, and obtaining from them a signed informed consent

² <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en>

form (see Annex 2: VICINITY consent form template), following the respective guidelines set in VICINITY and as described in section 3.4: Ethics and Security.

Beside this, certain guidelines will be implemented in order to limit the risk of data leaks;



- Keep anonymised data and personal data of respondents separate;
- Encrypt data if it is deemed necessary by the local researchers;
- Store data in at least two separate locations to avoid loss of data;
- Limit the use of USB flash drives;
- Save digital files in one of the preferred formats (see Annex 1), and
- Label files in a systematically structured way in order to ensure the coherence of the final dataset

A more formal description of best practice principles can be found in Table 1: Best practice for use of production data.

3.4. Production data

The consortium is aware that a number of privacy and data protection issues could be raised by the activities (use case demonstration and evaluation in WP7 and WP8) to be performed in the scope of the project. The project involves the carrying out of data collection in all pilot applications on the virtual neighborhood. For this reason, human participants will be involved in certain aspects of the system development by contributing real life data. During the development life cycle process, it will be necessary to operate on datasets. Some of the datasets may be based on production data, while others may be generated (synthetic).

The VICINITY architecture is decentralised by design (Figure 6). Production data will be used for testing purposes. Certain functionality like the discovery function and the related search criteria, raise the need for proper implementation of Things Ecosystem Description (TED) – which describes IoT assets that exists in the same environment.



Figure 6: The VICINITY architecture is decentralised by design

The public will have access to the VICINITY ontology alongside the VICINITY discovery function at the conclusion of the project. However, all data generated through the test phase and development process will be removed.

BEST PRACTICE – PRODUCTION DATA

The consortium will follow what is considered best practice for handling both copies of production data and live data.

- **Data Obfuscation and security safeguards**
Use obfuscation methods to remove/protect data or reduce the risk of personal information being harvested on data breach, and encrypt data where appropriate.
- **Data minimization**
Minimize the size of datasets and the amount of fields used.
- **Physical/environmental protection and access control**
Restrict and secure the environment where the data is used and stored and limit the ability to remove live data in either physical or electronic format from the environment. Also limit access to the data to authorized users with business needs and who have received appropriate data protection training.
- **Retention limits and data removal**
Limit the time period for use of the data and dispose of live data at end of use period. Destroy physical and electronic live data used for training, testing, or research at the conclusion of the project.
- **Use Limits**
Limit through controls and education the likelihood that live data, whose integrity is not reliable, is re-introduced into production systems or transferred to others beyond its intended purpose.
- **Watermarking**

<p>Include warning information on live data where possible to ensure users do not assume it is dummy data. This applies to all pilot sites where time critical actions have to be taken, and where forecast analysis needs to be based on accurate data.</p> <ul style="list-style-type: none"> Legal Controls Implement Confidentiality and Non-Disclosure Agreements if applicable. This will apply to all operators responsible for living labs that address eHealth and assisted living. Responsibility for accountability, training and awareness Ensure that identified personnel (by role) are assigned responsibility for compliance with any conditions of the approval for the use of live data. The personnel responsible for the technical description of the dataset will also serve as contact for the use of live data. This also applies to providing safety and training sessions for all persons having access to live data. The partners responsible for pilot sites handling real time data from living labs will prepare information that is to be handed out to relevant stakeholders.
--

Table 1: Best practice for use of production data

How these best practice principles are being implemented, are described in more detail in section 3.5 Ethics and Security and 3.11 Data sharing

3.5. Ethics and security

The consortium is aware that a number of privacy and data protection issues could be raised by the activities (use case demonstration and evaluation in WP7 and WP8) to be performed in the scope of the project. The project involves the carrying out of data collection in all pilot applications on the virtual neighborhood. For this reason, human participants will be involved in certain aspects of the project and data will be collected. This will be done in full compliance with any European and national legislation and directives relevant to the country where the data collections are taking place (INTERNATIONAL/EUROPEAN):

- The Universal Declaration of Human Rights and the Convention 108 for the Protection of Individuals with Regard to Automatic Processing of Personal Data and
- Directive 95/46/EC & Directive 2002/58/EC of the European parliament regarding issues with privacy and protection of personal data and the free movement of such data.

In addition to this, to further ensure that the fundamental human rights and privacy needs of participants are met whilst they take part in the project, in the Evaluation Plans a dedicated section will be delivered for providing ethical and privacy guidelines for the execution of the Industrial Trials. In order to protect the privacy rights of participants, a number of best practice principles will be followed. These include:

- no data will be collected without the explicit informed consent of the individuals under observation. This involves being open with participants about what they are involving themselves in and ensuring that they have agreed fully to the procedures/research being undertaken by giving their explicit consent.
- The owners of personal data are to be granted the right of inspection and the right to be removed from the registers.
- no data collected will be sold or used for any purposes other than the current project;
- a data minimisation policy will be adopted at all levels of the project and will be supervised by each Industrial Pilot Demonstration responsible. This will ensure that no data which is not strictly necessary to the completion of the current study will be collected;

- During the development life cycle process, it will be necessary to operate on datasets. Some of the datasets may be based on production data, while others may be generated (synthetic). These data will be removed by the end of the project.
- Any shadow (ancillary) personal data obtained during the course of the research will be immediately cancelled. However, the plan is to minimize this kind of ancillary data as much as possible. Special attention will also be paid to complying with the Council of Europe’s Recommendation R(87)15 on the processing of personal data for police purposes, Art.2 :

“The collection of data on individuals solely on the basis that they have a particular racial origin, particular religious convictions, sexual behavior or political opinions or belong to particular movements or organisations which are not proscribed by law should be prohibited. The collection of data concerning these factors may only be carried out if absolutely necessary for the purposes of a particular inquiry.”

- compensation – if and when provided – will correspond to a simple reimbursement for working hours lost as a result of participating in the study; special attention will be paid to avoid any form of unfair inducement;
- if employees of partner organizations, are to be recruited, specific measures will be in place in order to protect them from a breach of privacy/confidentiality and any potential discrimination; In particular their names will not be made public and their participation will not be communicated to their managers.
- Data should be pseudomised and anonymized to allow privacy to be upheld even if an attacker gains access to the system.
- Furthermore, if data has been compromised or tampering is detected, the involved parties are to be notified immediately in order to reduce risk of misuse of data gathered for research purposes.

The same concern addressed here also applies to open calls (see section 3.8 Open Call).

3.6. The VICINITY Data Management Portal

VICINITY will develop a data management portal as part of the project. This portal will provide to the public, for each dataset that will become publicly available, a description of the dataset along with a link to a download section. The portal will be updated each time a new dataset has been collected and is ready of public distribution. The portal will however not contain any datasets that should not become publicly available.

The initial version of the portal became available during the 2nd year of the project, in parallel to the establishment of the first versions of project datasets that can be made publicly available. The VICINITY data management portal will enable project partners to manage and distribute their public datasets through a common infrastructure as described in Table 2.

One dataset for (I/II)	One dataset for (II/II)	Administrative tools
each IoT unit	Datasets from pilots (see section 3.5 for examples)	List of sensor / grouping
personal information	groups of devices	List of actions / sequences
energy related domains	each health device	List of users
• each interface (energy)	node/object	List of contacts

• each measuring device (energy)	messaging	Balancing loads
• each routing device (energy)	sequences / actions (combination tokens / nodes)	Booking
mobility related domains	biometric (fingerprint, retina)	Messaging
• parking data (mobility)	camera	Criteria
• booking (mobility)	access	Priorities
• areas (mobility)	each smart home device (temperature, smoke, motion, sound)	Evaluation / feedback

Table 2: datasets stored in the VICINITY management portal

3.7. Format of datasets

For each dataset the following will be specified:

DS. PARTICiPANTName.##.Logical_sensorname	
Data Identification	
Dataset description	<i>Where are the sensor(s) installed? What are they monitoring/registering? What is the dataset comprised of? Will it contain future sub-datasets?</i>
Source (e.g. which device?)	<i>How will the dataset be collected? What kind of sensor is being used?</i>
Partners services and responsibilities	
Partner owner of the device	<i>What is the name of the owner of the device?</i>
Partner in charge of the data collection (if different)	<i>What is the name of the partner in charge of the device? Are there several partners that are cooperating? What are their names?</i>
Partner in charge of the data analysis (if different)	<i>The name of the partner.</i>
Partner in charge of the data storage (if different)	<i>The name of the partner.</i>
WPs and tasks	<i>The data are going to be collected within activities of WPxx and WPxx.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>What is the status with the metadata so far? Has it been defined? What is the content of the metadata (e.g. datatypes like images portraying an action, textual messages, sequences, timestamps etc.)</i>

Standards, Format, Estimated volume of data	<i>Has the data format been decided on yet? What will it look like?</i>
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Example text: Production process recognition and help during the different production phases, avoiding mistakes</i>
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<i>Example text: The full dataset will be confidential and only the members of the consortium will have access on it. Furthermore, if the dataset or specific portions of it (e.g. metadata, statistics, etc.) are decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section. Of course these data will be anonymized, so as not to have any potential ethical issues with their publication and dissemination</i>
Data sharing, re-use and distribution (How?)	<i>Has the data sharing policies been decided yet? What requirements exist for sharing data? How will the data be shared? Who will decide what to be shared?</i>
Embargo periods (if any)	-
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	<i>Who will own the information that has been collected? How will it adhere to partner policies? What kind of limitation is put on the archive?</i>

Table 3: Format of dataset description

3.8. Open Call

NB: The actual content of the Open Calls, such as documents and other material, is at the moment of this deliverable, still being worked out by the project participants. All descriptions and considerations related to open calls must therefore be considered tentative, and this section can be thought of as a tool for implementing best practise.

The Open Call process of the VICINITY project will involve third parties. System integrators (Figure 7) will be one of target groups for the calls. These will be presented for opportunities to integrate IoT infrastructures based on the VICINITY framework. Implementation/integration of Value-Added Services will also most likely be part of issues Open Calls will tackle. The calls should adhere to the principles which govern Commission calls. These principles all include confidentiality: all proposals and related data, knowledge and documents are treated in confidence.



Figure 7: Involving 3rd parties through open calls will provide VICINITY with valuable experience, and evolve interoperability

The Project Coordinator will present a legal contract with the third parties that are granted open calls. This contract will specify all the control procedures and made compliant with the Grant Agreement and the Consortium Agreement. This is done in order to assure that their contributions are in line with the agreed upon work plan; that the third party allows the Commission and the Court of Auditors to exercise their power of control on documents and information stored on electronic media or on the final recipient's premises.

Proposals for open calls and the deliverables that come as a result will include sections that describe how the data management principles have been implemented. It is expected the papers will follow the outlines that are presented in the legal contract and adhere to GDPR. This also applies to sharing ideas and intellectual properties. Furthermore, the deliverables will present how the chosen architecture and methodologies will be handled by the stakeholders, integrators and SME's.

According VICINITY concept the participants can decide with whom they wish to cooperate and to which extent. Participants will be held responsible for partners they team up with follow the same guidelines as the main project and the open call project.

3.9. Description of methods for dataset description

Example test dataset will be generated by research teams from the participants in the project. These test datasets will be prepared in XML-files. They will also be made available in XML and JSON format (Figure 8). The datasets will be based on semantic analysis of data from test sensors and applied to an ontology.

The collected dataset will encompass different methodological approaches and IoT standards defined by the global standard initiative oneM2M. The data will run through different test environments like TDD (Test Driven Development), ATDD (Acceptance Test Driven Development), PBT (Property Based Testing), BDD (Behavior Driven Development). The project will focus on using model-based test automation in processes with short release cycles.



Figure 8: Datasets will be prepared and provided in XML and JSON format

Apart from the research teams, these datasets will be useful for other research groups, Standard Development Organisations (SDO) and technical integrators with within the area of Internet of Things (IoT).

No comparable data is available as of yet, but there are several descriptions that will be used as basis for the test data.

All datasets are to be shared between the participants during the lifecycle of the project. Feedback from other participants and test implementations will decide when the dataset should be made publicly available. When the datasets support the framework defined by the VICINITY ontology, they will be made public and presented in open access publications.

The VICINITY partners can use a variety of methods for exploitation and dissemination of the data including:

- Using them in further research activities (outside the action)
- Developing, creating or marketing a product or process
- Creating and providing a service, or
- Using the data in standardisation activities

Restrictions:

- 1) All national reports (which include data and information on the relevant topic) will be available to the public through the HERON web-site or a repository or any other option that the consortium decides and after verification by the partners so as to ensure their quality and credibility.
- 2) After month 18 so that partners have the time to produce papers; 3) Open access to the research data itself is not applicable.

3.10. Standards and metadata

The data will be generated and tested through different test automation technologies, e.g. TDL (Test description language), TTCN-3 (Test and Test Control Notation), UTP (UML Testing Profile). The profile should mimic the data communicated from IoT units following the oneM2M specifications. The Systems Modeling Language³ (SysML) is used for the collection, analysis and processing of requirements as well as for the specification message exchanges and overviews of architecture and behavior specifications (Figure 9).

³ www.omg.sysml.org

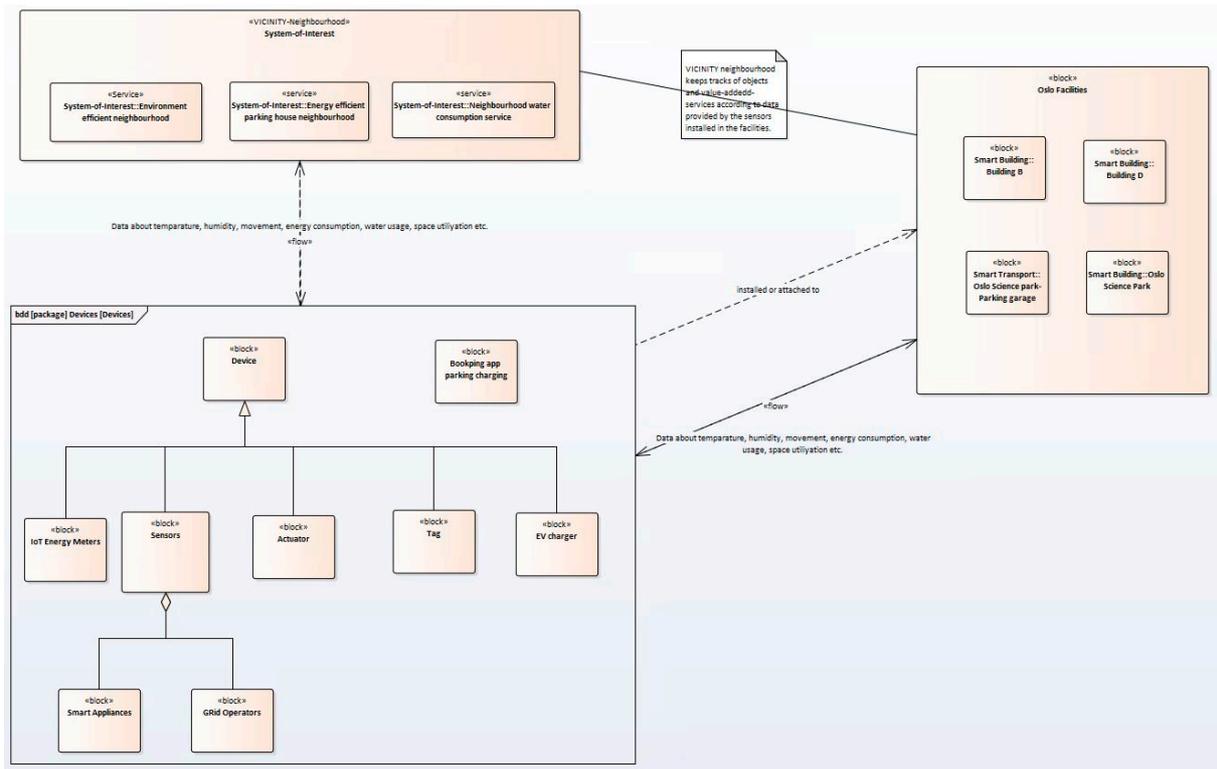


Figure 9: Example of SysML model of Virtual Oslo Science City

The project intends to share the datasets in an internally accessible disciplinary repository using descriptive metadata as required/provided by that repository. Additional metadata to example test datasets will be offered within separate XML-files.

They will also be made available in XML and JSON format. Keywords will be added as notations in SysML and modelled on the specifications defined by oneM2M. The content will be similar to relevant data from compatible IoT devices and network protocols. No network protocols have been defined yet, but several have been evaluated. Files and folders will be versioned and structured by using a name convention consisting of project name, dataset name, date, version and ID.

3.11. Data sharing

The project aims to prepare the API for internal testing through the VICINITY open gateway.

The VICINITY open gateway is defined as Interoperability as a Service. In other words - it is a cloud based service that assumes the data has already been gathered and transferred to the software running on the service layer. These data will be made available for researchers in a controlled environment, where login credentials are used to get access to the data in XML and JSON-format (Figure 10).



Figure 10: Data will only be provided partners with proper login credentials

The project focus on developing a framework that allows for a scalable and futureproof platform upon which it can invest and develop IoT applications, without fear of vendor lock-in or needing to commit to one connectivity technology.

The researchers must therefore be committed to the requirements, architecture, application programming interface (API) specifications, security solutions and mapping to common industry protocols such as CoAP, MQTT and HTTP. Further analysis will be performed using freely available open source software tools. The data will also be made available as separate files.

The goal is to ultimately support the Europe 2020 strategy⁴ by offering the open data portal. The Digital Agenda proposes to better exploit the potential of Information and Communication Technologies (ICTs) in order to foster innovation, economic growth and progress. Thus VICINITY will support EUs efforts in exploiting the potential offered by using ICT in areas like climate change, managing ageing population, and intelligent transport system to mention a few examples.

3.12. Archiving and preservation (including storage and backup)

As specified by the "rules of good scientific practice" we aim to preserve data for at least ten years. Approximated end volume of example test dataset is currently 10 GB, but this may be subject to change as the scope of the project may change.

Associated costs for dataset preparation for archiving will be covered by the project itself, while long term preservation will be provided and associated costs covered by a selected disciplinary repository. During the project data will be stored on the VICINITY web cloud as well as being replicated to a separate external server.

⁴ <https://ec.europa.eu/digital-single-market/en/europe-2020-strategy>

4. Datasets for smart grid from Aalborg University (AAU)



AALBORG UNIVERSITY
DENMARK

AAU will mainly deal with control design, energy management systems implementation and Information and Communication Technology (ICT) integration in small scale energy systems. AAU will scale-up by using hardware in the loop solution and will participate actively in the implementation at the Energy sites proposed in VICINITY. AAU will act as interface between ICT experts and Energy sites in the project, as well as test interactions between the developed concepts on the ICT side and the control and management of electric power networks. Implementation and experimental results will be an important outcome for the project.

DS.AAU.01.GRID_Status	
Data Identification	
Dataset description	<i>This dataset comprised different parameters characterising the electrical grid from the generation to the distribution sections. The cost of the electricity will also be considered in this dataset, so as to have full information that enables micro-trading actions.</i>
Source (e.g. which device?)	<i>The sensors that feed this dataset are; energy generation and consumption on-site from RES, instant grid cost of energy consumed and purchased from the grid</i>
Partners services and responsibilities	
Partner owner of the device	<i>The devices will be the property of the test site owners, where the data collection is going to be performed.</i>
Partner in charge of the data collection (if different)	AAU
Partner in charge of the data analysis (if different)	AAU
Partner in charge of the data storage (if different)	AAU
WPs and tasks	<i>The data are going to be collected within activities of WP7 and WP8.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.</i>
Standards, Format, Estimated volume of data	<i>The data are received in JSON format. Regarding the volume of data, it depends on the motion/activity levels of the engaged devices. However, it is estimated to be 4 KB/transmission.</i>



Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Due to privacy issues, the collected data are stored at a secured database scheme at Aalborg University Facilities. Data exploitation is foreseen to be achieved through testing value-added services, data analytics and statistical analysis.</i>
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<i>The full dataset will be confidential and only the authorized AAU personnel will have access as defined. AAU could provide energy data to specific consortium members under a detailed confidentiality framework.</i>
Data sharing, re-use and distribution (How?)	<i>The created dataset could be shared under a detailed confidentiality framework by using open APIs through the middleware.</i>
Embargo periods (if any)	<i>None</i>
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	<i>Due to privacy issues, the collected data are stored at a secured database scheme at Aalborg University Facilities. Data exploitation is foreseen to be achieved through testing value-added services, data analytics and statistical analysis. A back up will be stored in an external storage device, kept by AAU in a secured place. This back-up will be available when it is required from the pilot sites.</i>

Table 4: Dataset description of the AAU GRID status

5. Datasets for smart energy from Enercoutim (ENERC)



ENERC will participate providing the facilities and the experience in implementing solar production integrated into municipality smart city efforts. To this end, ENERC will actively participate in the deployment, management and evaluation of the “Smart Energy Microgrid Neighbourhood” Use Case. Its contribution will be focused on the energy resource potential demand studies and economic sustainability. Its expertise will allow ICT integration with smart city management focused on better serving its citizens.

The main aim of this project is the demonstration of a Solar Platform which provides a set of shared infrastructures and reduces the total cost per MW as well as improves the environmental impact compared to the stand alone implementation of these projects. As main responsibilities, ENERC will be in charge of strategic technology planning and integration coordination, designing potential models for municipal energy management, as well as identifying the optimal ownership structure of the microgrid system with a focus on delivering maximum social and economic benefit to the local community.

DS.ENERC.01.METEO_Station	
Data Identification	
Dataset description	<i>The weather conditions will influence the energy production, so it becomes critical to understand the current and foreseen scenarios. It is fundamental to constantly carry out different measures with the meteo station equipment of the parameters that can influence both energy production and consumption over time.</i>
Source (e.g. which device?)	<i>The sensors that feed this dataset are; temperature, humidity, wind speed and wind direction, barometer, precipitation measurement and sun tracker.</i>
Partners services and responsibilities	
Partner owner of the device	<i>The devices will be the property of the test site owners, where the data collection is going to be performed.</i>
Partner in charge of the data collection (if different)	<i>ENERC</i>
Partner in charge of the data analysis (if different)	<i>ENERC</i>
Partner in charge of the data storage (if different)	<i>ENERC</i>
WPs and tasks	<i>The data are going to be collected within activities of WP7 and WP8.</i>
Standards	

Info about metadata (Production and storage dates, places) and documentation?	<i>The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.</i>
Standards, Format, Estimated volume of data	<i>The data are received in JSON format. Regarding the volume of data, it depends on the motion/activity levels of the engaged devices. However, it is estimated to be 4 KB/transmission.</i>
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Due to privacy issues, the collected data are stored at a secured database scheme at SOLAR LAB Facilities, allowing access to registered users. Data exploitation is foreseen to be extended through envisioned value-added services, allowing full access to specific authorised users (e.g. facility managers), and for a broader use in an anonymised/aggregated manner for data analytics and statistical analysis.</i>
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<i>The full dataset will be confidential and only the authorized ENERC personnel and related end-users will have access as defined. Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework. Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.</i>
Data sharing, re-use and distribution (How?)	<i>The created dataset could be shared by using open APIs through the middleware as well as a data management portal.</i>
Embargo periods (if any)	<i>None</i>
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	<i>Due to ethical and privacy issues, data will be stored in a database scheme at the SOLAR LAB facilities, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by ENERC in a secured place. Data will be kept indefinitely allowing statistical analysis.</i>

Table 5: Dataset description of the ENERC METEO station

DS.ENERC.02.BUILDING_Status	
Data Identification	
Dataset description	<i>The information associated to the energy consumption in buildings will allow identifying the usage of resources for each measurement point.</i>
Source (e.g. which device?)	<i>The sensors that feed this dataset are; Cooling energy demand, heating energy demand, hot water demand, building equipment demand</i>
Partners services and responsibilities	

Partner owner of the device	<i>The devices will be the property of the test site owners, where the data collection is going to be performed.</i>
Partner in charge of the data collection (if different)	<i>ENERC</i>
Partner in charge of the data analysis (if different)	<i>ENERC</i>
Partner in charge of the data storage (if different)	<i>ENERC</i>
WPs and tasks	<i>The data are going to be collected within activities of WP7 and WP8.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.</i>
Standards, Format, Estimated volume of data	<i>The data are received in JSON format. Regarding the volume of data, it depends on the motion/activity levels of the engaged devices. However, it is estimated to be 4 KB/transmission.</i>
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Due to privacy issues, the collected data are stored at a secured database scheme at SOLAR LAB Facilities, allowing access to registered users. Data exploitation is foreseen to be extended through envisioned value-added services, allowing full access to specific authorised users (e.g. facility managers), and for a broader use in an anonymised/aggregated manner for data analytics and statistical analysis.</i>
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<i>The full dataset will be confidential and only the authorized ENERC personnel and related end-users will have access as defined. Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework. Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.</i>
Data sharing, re-use and distribution (How?)	<i>The created dataset could be shared by using open APIs through the middleware as well as a data management portal.</i>
Embargo periods (if any)	<i>None.</i>
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	<i>Due to ethical and privacy issues, data will be stored in a database scheme at the SOLAR LAB facilities, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by ENERC in a secured place. Data will be kept indefinitely allowing statistical analysis.</i>

Table 6: Dataset description of the ENERC building status

DS.ENERC.03.GRID_Status

Data Identification	
Dataset description	<i>This dataset comprises the different parameters that characterise the electrical grid from the generation to the distribution sections. Moreover the cost of the electricity will be considered in this dataset so as to have full information that enables micro-trading actions.</i>
Source (e.g. which device?)	<i>The sensors that feed this dataset are; Electrical energy generated on-site from RES ,Thermal energy generated on-site, thermal energy consumed, grid electricity consumed, instant grid cost of energy consumed, value of energy purchased from the grid</i>
Partners services and responsibilities	
Partner owner of the device	<i>The devices will be the property of the test site owners, where the data collection is going to be performed.</i>
Partner in charge of the data collection (if different)	<i>ENERC, AAU</i>
Partner in charge of the data analysis (if different)	<i>ENERC, AAU</i>
Partner in charge of the data storage (if different)	<i>ENERC, AAU</i>
WPs and tasks	<i>The data are going to be collected within activities of WP7 and WP8.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.</i>
Standards, Format, Estimated volume of data	<i>The data are received in JSON format. Regarding the volume of data, it depends on the motion/activity levels of the engaged devices. However, it is estimated to be 4 KB/transmission.</i>
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Due to privacy issues, the collected data are stored at a secured database scheme at SOLAR LAB Facilities and AAU servers, allowing access to registered users. Data exploitation is foreseen to be extended through envisioned value-added services, allowing full access to specific authorised users (e.g. facility managers), and for a broader use in an anonymised/aggregated manner for data analytics and statistical analysis.</i>

<p>Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public</p>	<p><i>The full dataset will be confidential and only the authorized ENERC/AAU personnel and related end-users will have access as defined. Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework.</i></p> <p><i>Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.</i></p>
<p>Data sharing, re-use and distribution (How?)</p>	<p><i>The created dataset could be shared by using open APIs through the middleware as well as a data management portal.</i></p>
<p>Embargo periods (if any)</p>	<p><i>None</i></p>
<p>Archiving and preservation (including storage and backup)</p>	
<p>Data storage (including backup): where? For how long?</p>	<p><i>Due to ethical and privacy issues, data will be stored in a database scheme at the SOLAR LAB facilities and AAU servers, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by ENERC in a secured place. Data will be kept indefinitely allowing statistical analysis.</i></p>

Table 7: Dataset description of the ENERC grid status

6. Datasets for eHealth from GNOMON Informatics SA (GNOMON)



GNOMON will provide its background knowledge in the specific field of assisted living and tele care in the context of social workers. In addition, GNOMON will actively contribute in the use case pilot setup, assessment and benchmarking.

The company has developed and provided the remote care and monitoring integrated system for people with health problems as well as of the software applications for support and organization using information and communication technologies of the business operation of HELP AT HOME program in the Municipality of Pilea-Hortiatis. This infrastructure could be further exploited and extended for the scope of VICINITY project and specifically for the realisation of the eHealth Use Case.

DS.GNOMON.01.Pressure_sensor

Data Identification	
Dataset description	<i>The sensors will be in possession of patients in need of assisted living and identified by the equivalent municipality (MPH) health care services to ensure the validity of each case. The measurements are scheduled to be taken once a day, requiring the patient to make use of the device placed within their apartment. The main task of the sensor is to monitor pressure (systolic/diastolic) and heart rate levels.</i>
Source (e.g. which device?)	<i>The dataset will be collected via a combination of connected devices consisting of a Bluetooth Blood Pressure monitor and a Connectivity Gateway based on Raspberry pi.</i>
Partners services and responsibilities	
Partner owner of the device	<i>The device will be the property of the test site owners, where the data collection is going to be performed.</i>
Partner in charge of the data collection (if different)	<i>GNOMON, MPH</i>
Partner in charge of the data analysis (if different)	<i>GNOMON, MPH</i>
Partner in charge of the data storage (if different)	<i>GNOMON, MPH</i>
WPs and tasks	<i>The data are going to be collected within activities of WP7 and WP8.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.</i>

Standards, Format, Estimated volume of data	<i>The data are received in XML format. In a later stage, they are converted to JSON format and stored in a database. Regarding the volume of data, it depends on the participation levels of the engaged patients. However, it is estimated to be 16 KB/measurement.</i>
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Due to privacy issues, the collected data are stored at a secured database scheme at MPH headquarters, allowing access to registered users (i.e. MPH health care services personnel and eHealth call center). Data exploitation is foreseen to be extended through envisioned value-added services, allowing full access to specific authorised users (e.g. doctors), and for a broader use in an anonymised/aggregated manner for creating behaviour profiles and clustering patients to different medical groups.</i>
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<i>The full dataset will be confidential and only the authorized MPH personnel and related end-users will have access as defined. The latter authorized groups of users will access data in a tamper-proof way with an audit mechanism triggered simultaneously to guarantee the alignment with relevant requirements coming from the recently introduced General Data Protection Regulation (GDPR). Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework.</i> <i>Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.</i>
Data sharing, re-use and distribution (How?)	<i>The created dataset could be shared by using open APIs through the middleware as well as a data management portal. Dataset from VICINITY could be used and exploited anonymized from another European project. Dataset from health devices deployed at seniors' houses will provide added value and be the base for other research projects (e.g. statistical data). VICINITY could have an open portal / repository on its website, providing anonymized data's information like timestamp and description.</i>
Embargo periods (if any)	<i>None</i>
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	<i>Due to ethical and privacy issues, data will be stored in a database scheme at the headquarters of MPH, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by MPH in a secured place. Data will be kept indefinitely allowing statistical analysis.</i>

Table 8: Dataset description of the GNOMON pressure sensor

DS.GNOMON.02.Weight_sensor

Data Identification	
Dataset description	<i>The sensors will be in possession of patients in need of assisted living and identified by the equivalent municipality (MPH) health care services to ensure the validity of each case. The measurements are scheduled to be taken once a day, requiring the patient to make use of the device placed within their apartment. The main task of the sensor is to keep track of weight measurements and mass index (given the fact that the patient provides an accurate value of his/her height). Future subset may contain information about resting metabolism, visceral fat level, skeletal muscle and body age.</i>
Source (e.g. which device?)	<i>The dataset will be collected via a combination of connected devices consisting of a Bluetooth Body Composition monitor and a Connectivity Gateway based on Raspberry pi.</i>
Partners services and responsibilities	
Partner owner of the device	<i>The device will be the property of the test site owners, where the data collection is going to be performed.</i>
Partner in charge of the data collection (if different)	<i>GNOMON, MPH</i>
Partner in charge of the data analysis (if different)	<i>GNOMON, MPH</i>
Partner in charge of the data storage (if different)	<i>GNOMON, MPH</i>
WPs and tasks	<i>The data are going to be collected within activities of WP7 and WP8.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.</i>
Standards, Format, Estimated volume of data	<i>The data are received in XML format. In a later stage, they are converted to JSON format and stored in a database. Regarding the volume of data, it depends on the participation levels of the engaged patients. However, it is estimated to be 48 KB/measurement.</i>
Data exploitation and sharing	

Data exploitation (purpose/use of the data analysis)	<i>Due to privacy issues, the collected data are stored at a secured database scheme at MPH headquarters, allowing access to registered users (i.e. MPH health care services personnel and eHealth call center). Data exploitation is foreseen to be extended through envisioned value-added services, allowing full access to specific authorised users (e.g. doctors), and for a broader use in an anonymised/aggregated manner for creating behaviour profiles and clustering patients to different medical groups.</i>
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<i>The full dataset will be confidential and only the authorized MPH personnel and related end-users will have access as defined. The latter authorized groups of users will access data in a tamper-proof way with an audit mechanism triggered simultaneously to guarantee the alignment with relevant requirements coming from the recently introduced General Data Protection Regulation (GDPR). Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework.</i> <i>Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.</i>
Data sharing, re-use and distribution (How?)	<i>The created dataset could be shared by using open APIs through the middleware as well as a data management portal. Dataset from VICINITY could be used and exploited anonymized from another European project. Dataset from health devices deployed at seniors' houses will provide added value and be the base for other research projects (e.g. statistical data). VICINITY could have an open portal / repository on its website, providing anonymized data's information like timestamp and description.</i>
Embargo periods (if any)	<i>None</i>
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	<i>Due to ethical and privacy issues, data will be stored in a database scheme at the headquarters of MPH, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by MPH in a secured place. Data will be kept indefinitely allowing statistical analysis.</i>

Table 9: Dataset description of the GNOMON weight sensor

DS.GNOMON.03.Fall_sensor

Data Identification

Dataset description	<i>The fall sensor is a wearable sensor that will be in possession of patients in need of assisted living and identified by the equivalent municipality (MPH) health care services to ensure the validity of each case. The main goal of the sensor is to automatically detect when a patient falls either due to an accident or in the case of a medical incident. The event is triggered automatically after a fall, but a similar event is also triggered by pressing the equivalent panic button (wearable actuator). In both cases, an automated emergency phone call is placed to the eHealth Call Center.</i>
Source (e.g. which device?)	<i>The dataset will be collected via a combination of devices consisting of a hub (Lifeline Vi) and a fall detector that are wirelessly connected.</i>
Partners services and responsibilities	
Partner owner of the device	<i>The device will be the property of the test site owners, where the data collection is going to be performed.</i>
Partner in charge of the data collection (if different)	<i>GNOMON, MPH</i>
Partner in charge of the data analysis (if different)	<i>GNOMON, MPH</i>
Partner in charge of the data storage (if different)	<i>GNOMON, MPH</i>
WPs and tasks	<i>The data are going to be collected within activities of WP7 and WP8.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.</i>
Standards, Format, Estimated volume of data	<i>An audit log containing alerts (incl. false alarms) is stored. The amount of alerts is estimated to be 50 alerts (incl. false alarms) per month.</i>
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Due to privacy issues, the collected data are stored at a secured database scheme at MPH headquarters, allowing access to registered users (i.e. MPH health care services personnel and eHealth call centre). Data exploitation is foreseen to be extended through envisioned value-added services, allowing full access to specific authorised users (e.g. patient's doctors), and for a broader use in an anonymised/aggregated manner for data analytics and statistical analysis.</i>

Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<p><i>The full dataset will be confidential and only the authorized MPH personnel and related end-users will have access as defined. The latter authorized groups of users will access data in a tamper-proof way with an audit mechanism triggered simultaneously to guarantee the alignment with relevant requirements coming from the recently introduced General Data Protection Regulation (GDPR). Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework.</i></p> <p><i>Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.</i></p>
Data sharing, re-use and distribution (How?)	<p><i>The created dataset could be shared by using open APIs through the middleware as well as a data management portal. Dataset from VICINITY could be used and exploited anonymized from another European project. Dataset from fall sensors at seniors' houses will provide added value and be the base for other research projects (e.g. statistical data). VICINITY could have an open portal / repository on its website, providing anonymized data's information like timestamp and description.</i></p>
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	<p><i>Due to ethical and privacy issues, data will be stored in a database scheme at the headquarters of MPH, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by MPH in a secured place. Data will be kept indefinitely allowing statistical analysis.</i></p>

Table 10: Dataset description of the GNOMON fall sensor

DS.GNOMON.04.Wearable_Fitness_Tracker_Sensor

Data Identification	
Dataset description	<p><i>The fitness sensors are sensors embodied to wearable fitness trackers such as activity wristbands. The latter equipment will be in possession of middle aged citizens, either with a chronic health issue (e.g. obesity) or not, that are identified by the equivalent municipality (MPH). The municipality will try to promote fitness awareness and improve citizens' health under the concept of a municipal-scale competition that will be based on activity related data coming from the sensors (e.g. step counting, hours of sleep, etc).</i></p>
Source (e.g. which device?)	<p><i>The data will be collected by wearable fitness trackers, mainly in the form of activity wristbands (e.g. Xiaomi MiBand, FitBit, etc.).</i></p>
Partners services and responsibilities	

Partner owner of the device	<i>The device will be the property of the test subject, in this case the participating citizen.</i>
Partner in charge of the data collection (if different)	<i>GNOMON, MPH</i>
Partner in charge of the data analysis (if different)	<i>GNOMON, MPH</i>
Partner in charge of the data storage (if different)	<i>GNOMON, MPH</i>
WPs and tasks	<i>The data are going to be collected within activities of WP7 and WP8.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.</i>
Standards, Format, Estimated volume of data	<i>The collection of data from wearable fitness tracker sensors is event-driven. New data are dispatched once they are produced.</i>
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Data exploitation is foreseen to be extended through envisioned value-added services, allowing full access to specific authorised users (e.g. doctors), and for a broader use in an anonymised/aggregated manner for data analytics and statistical analysis. Additionally, as one of the value-added services introduced is related to the concept of a municipal-scale competition, data analysis will also serve the needs of calculating and providing a ranking among the competitors.</i>
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<p><i>The full dataset will be confidential and only the authorized MPH personnel and related end-users will have access as defined. The latter authorized groups of users will access data in a tamper-proof way with an audit mechanism triggered simultaneously to guarantee the alignment with relevant requirements coming from the recently introduced General Data Protection Regulation (GDPR). Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework.</i></p> <p><i>Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.</i></p>

Data sharing, re-use and distribution (How?)	<i>The created dataset could be shared by using open APIs through the middleware as well as a data management portal. Dataset from VICINITY could be used and exploited anonymized from another European project. Dataset from wearable fitness trackers will provide added value and be the base for other research projects (e.g. statistical data). VICINITY could have an open portal / repository on its website, providing anonymized data's information like timestamp and description.</i>
Embargo periods (if any)	<i>None</i>
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	<i>Due to ethical and privacy issues, data will be stored in a database scheme at the headquarters of MPH, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by MPH in a secured place. Data will be kept indefinitely allowing statistical analysis.</i>

Table 11: Dataset description of the GNOMON Wearable Fitness Tracker Sensor

DS.GNOMON.05.Beacon_Sensor

Data Identification	
Dataset description	<i>The beacon sensors are sensors to be deployed in municipality's sport facilities, e.g. gym, pool, etc. and also tested at CERTH/ITI's Smart Home. The municipality will try to promote fitness awareness and improve citizens' health under the concept of a municipal-scale competition that will be based on activity related data gathered by the sensors and processed accordingly (e.g. translation of beacon signals to actual time spent in sport facilities).</i>
Source (e.g. which device?)	<i>The data will be collected by beacons deployed in municipality's sport facilities and at CERTH/ITI's Smart Home.</i>
Partners services and responsibilities	
Partner owner of the device	<i>The device will be the property of the test site owners, where the data collection is going to be performed.</i>
Partner in charge of the data collection (if different)	<i>GNOMON, MPH, CERTH</i>
Partner in charge of the data analysis (if different)	<i>GNOMON, MPH, CERTH</i>
Partner in charge of the data storage (if different)	<i>GNOMON, MPH, CERTH</i>
WPs and tasks	<i>The data are going to be collected within activities of WP7 and WP8.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.</i>

Standards, Format, Estimated volume of data	<i>The collection of data from beacons is event-driven. New data are dispatched once they are produced for example when middle-age person visits a sport centre.</i>
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Data exploitation is foreseen to be extended through envisioned value-added services, allowing full access to specific authorised users (e.g. doctors), and for a broader use in an anonymised/aggregated manner for data analytics and statistical analysis. Additionally, as one of the value-added services introduced is related to the concept of a municipal-scale competition, data analysis will also serve the needs of calculating and providing a ranking among the competitors.</i>
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<p><i>The full dataset of beacons deployed in CERTH / ITI's smart house, that is not sensitive, will be accessible through a local experimental repository.</i></p> <p><i>The full dataset of beacons deployed in houses of elderly people are sensitive, therefore, will be confidential and only the authorized MPH personnel and related end-users will have access as defined. The latter authorized groups of users will access data in a tamper-proof way with an audit mechanism triggered simultaneously to guarantee the alignment with relevant requirements coming from the recently introduced General Data Protection Regulation (GDPR). Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework.</i></p> <p><i>Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.</i></p>
Data sharing, re-use and distribution (How?)	<i>The created dataset could be shared by using open APIs through the middleware as well as a data management portal. Dataset from VICINITY could be used and exploited anonymized from another European project. Dataset from beacons at sport centres will provide added value and be the base for other research projects (e.g. statistical data). VICINITY could have an open portal / repository on its website, providing anonymized data's information like timestamp and description.</i>
Embargo periods (if any)	<i>None</i>
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	<i>Due to ethical and privacy issues, data will be stored in a database scheme at the headquarters of MPH, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by MPH in a secured place. Data will be kept indefinitely allowing statistical analysis.</i>

Table 12: Dataset description of the GNOMON Beacon Sensor

DS.GNOMON_CERTH.06.Gorenje_Smart_Appliances_Sensor

Data Identification	
Dataset description	<i>The sensors related to Gorenje smart appliances are sensors embodied to specific house equipment such as ovens and fridges. The latter equipment will be provided by Gorenje partner and will be in possession of patients in need of assisted living and identified by the equivalent municipality (MPH) health care services to ensure the validity of each case. Similar equipment will also be deployed in CERTH / ITI's facilities. The main goal of the sensors is to automatically detect when a patient opens the fridge or uses the oven in order to create behaviour profiles based on relevant criteria (e.g. frequency of use, etc), trigger alerts in case of deviation from the normal standards of use and inform the call centre.</i>
Source (e.g. which device?)	<i>The data will be collected by specific smart appliances (i.e. oven, fridge) provided by the Gorenje partner and adjusted to VICINITY requirements.</i>
Partners services and responsibilities	
Partner owner of the device	<i>The device will be the property of the test site owners, where the data collection is going to be performed.</i>
Partner in charge of the data collection (if different)	<i>CERTH, GORENJE, GNOMON, MPH</i>
Partner in charge of the data analysis (if different)	<i>CERTH, GORENJE, GNOMON, MPH</i>
Partner in charge of the data storage (if different)	<i>CERTH, GORENJE, GNOMON, MPH</i>
WPs and tasks	<i>The data are going to be collected within activities of WP6, WP7 and WP8.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.</i>
Standards, Format, Estimated volume of data	<i>The collection of data from Gorenje devices is time-driven and dispatched every 15min and it is depended on the standards that Gorenje provides.</i>
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Data exploitation is foreseen to be extended through envisioned value-added services and for a broader use in an anonymised/aggregated manner for creating behaviour profiles and clustering patients to different medical groups. Significant deviation from the latter profiles is expected to trigger relevant alerts.</i>

<p>Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public</p>	<p><i>The full dataset of Gorenje devices deployed in CERTH / ITI's facilities, that are not sensitive, will be accessible through Gorenje Cloud in a local experimental repository.</i></p> <p><i>The full dataset from Gorenje devices deployed in elderly's people houses will be confidential and only the authorized MPH personnel and related end-users will have access as defined through Gorenje Cloud. The latter authorized groups of users will access data in a tamper-proof way with an audit mechanism triggered simultaneously to guarantee the alignment with relevant requirements coming from the recently introduced General Data Protection Regulation (GDPR). Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework.</i></p> <p><i>Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.</i></p>
<p>Data sharing, re-use and distribution (How?)</p>	<p><i>The created dataset could be shared by using open APIs through the middleware as well as a data management portal. Dataset from VICINITY could be used and exploited anonymized from another European project. Dataset from Gorenje devices deployed at seniors' houses will provide added value and be the base for other research projects (e.g. statistical data). VICINITY could have an open portal / repository on its website, providing anonymized data's information like timestamp and description.</i></p>
<p>Embargo periods (if any)</p>	<p><i>None</i></p>
<p>Archiving and preservation (including storage and backup)</p>	
<p>Data storage (including backup): where? For how long?</p>	<p><i>Due to ethical and privacy issues, data will be stored in a database scheme at the headquarters of MPH, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by MPH in a secured place. Data will be kept indefinitely allowing statistical analysis.</i></p>

Table 13: Dataset description of the GNOMON/CERTH Gorenje Smart Appliances Sensor

7. Datasets for eHealth from Centre for Research and Technology Hellas (CERTH)



CERTH / ITI will contribute in the use case pilot setup for houses at Municipality of Pilea-Hortiatis and provide its background knowledge in the field of assisted living. It will also provide its infrastructure of Smart House for cross-domain implementation including building sensors and devices which have been also integrated to houses at MPH.

DS.CERTH.01.Occupancy_Sensor

Data Identification	
Dataset description	<i>Occupancy sensors will be deployed, on the one hand, in houses of patients in need of assisted living, identified by the equivalent municipality (MPH) health care services to ensure the validity of each case, but also in CERTH's smart house facilities for testing reasons. The main task of the sensor is to provide a 24/7 occupancy status for the area of its responsibility. Data coming from this sensor will be used to create behaviour profiles based on relevant criteria (e.g. occupancy level for a specific room, etc) and trigger alerts in case of deviation from the normal standards.</i>
Source (e.g. which device?)	<i>The dataset will be collected via a combination of connected occupancy sensors (e.g. Wi-Fi, ZigBee etc.) and a Connectivity Gateway based on Raspberry pi or other vendor.</i>
Partners services and responsibilities	
Partner owner of the device	<i>The device will be the property of the test site owners, where the data collection is going to be performed.</i>
Partner in charge of the data collection (if different)	<i>GNOMON, MPH, CERTH</i>
Partner in charge of the data analysis (if different)	<i>GNOMON, MPH, CERTH</i>
Partner in charge of the data storage (if different)	<i>GNOMON, MPH, CERTH</i>
WPs and tasks	<i>The data are going to be collected within activities of WP7 and WP8.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.</i>
Standards, Format, Estimated volume of data	<i>The collection of data from occupancy sensors is time-driven and dispatched every 15min (e.g. through REST Services, XML format etc.).</i>

Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Data exploitation is foreseen to be extended through envisioned value-added services and for a broader use in an anonymised/aggregated manner for creating behaviour profiles and clustering patients to different medical groups. Significant deviation from the latter profiles is expected to trigger relevant alerts which will be sent to the call centre.</i>
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<p><i>The full dataset of occupancy sensors deployed in CERTH / ITI's smart house, that are not sensitive, will be accessible through a local experimental repository.</i></p> <p><i>The full dataset of sensors deployed in houses of elderly people are sensitive therefore will be confidential and only the authorized MPH personnel and related end-users will have access as defined. The latter authorized groups of users will access data in a tamper-proof way with an audit mechanism triggered simultaneously to guarantee the alignment with relevant requirements coming from the recently introduced General Data Protection Regulation (GDPR). Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework.</i></p> <p><i>Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.</i></p>
Data sharing, re-use and distribution (How?)	<i>The created dataset could be shared by using open APIs through the middleware as well as a data management portal. Dataset from VICINITY could be used and exploited anonymized from another European project. Dataset from sensors deployed at seniors' houses will provide added value and be the base for other research projects (e.g. statistical data). VICINITY could have an open portal / repository on its website, providing anonymized data's information like timestamp and description.</i>
Embargo periods (if any)	<i>None</i>
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	<i>Due to ethical and privacy issues, data will be stored in a database scheme at the headquarters of MPH, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by MPH in a secured place. Data will be kept indefinitely allowing statistical analysis.</i>

Table 14: Dataset description of the CERTH Occupancy Sensor

DS.CERTH.02.Motion_Sensor

Data Identification

Dataset description	<i>Motion sensors will be deployed, on the one hand, in houses of patients in need of assisted living, identified by the equivalent municipality (MPH) health care services to ensure the validity of each case, but also in CERTH's smart house facilities for testing reasons. The main task of the sensor is to provide the 24/7 motion levels for the area of its responsibility. Data coming from this sensor will be used to create behaviour profiles based on relevant criteria (e.g. motions level for a specific room and time period, etc.) and trigger alerts in case of deviation from the normal standards.</i>
Source (e.g. which device?)	<i>The dataset will be collected via a combination of connected motion sensors (e.g. Wi-Fi, ZigBee etc.) and a Connectivity Gateway based on Raspberry pi or other vendor.</i>
Partners services and responsibilities	
Partner owner of the device	<i>The device will be the property of the test site owners, where the data collection is going to be performed.</i>
Partner in charge of the data collection (if different)	<i>GNOMON, MPH, CERTH</i>
Partner in charge of the data analysis (if different)	<i>GNOMON, MPH, CERTH</i>
Partner in charge of the data storage (if different)	<i>GNOMON, MPH, CERTH</i>
WPs and tasks	<i>The data are going to be collected within activities of WP6, WP7 and WP8.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.</i>
Standards, Format, Estimated volume of data	<i>The collection of data from motion sensors is time-driven and dispatched every 15min (e.g. through REST Services, XML format etc.).</i>
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Data exploitation is foreseen to be extended through envisioned value-added services and for a broader use in an anonymised/aggregated manner for creating behaviour profiles and clustering patients to different medical groups. Significant deviation from the latter profiles is expected to trigger relevant alerts.</i>

<p>Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public</p>	<p><i>The full dataset of occupancy sensors deployed in CERTH / ITI's smart house, that is not sensitive, will be accessible through a local experimental repository.</i></p> <p><i>The full dataset of sensors deployed in houses of elderly people are sensitive therefore will be confidential and only the authorized MPH personnel and related end-users will have access as defined. The latter authorized groups of users will access data in a tamper-proof way with an audit mechanism triggered simultaneously to guarantee the alignment with relevant requirements coming from the recently introduced General Data Protection Regulation (GDPR). Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework.</i></p> <p><i>Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.</i></p>
<p>Data sharing, re-use and distribution (How?)</p>	<p><i>The created dataset could be shared by using open APIs through the middleware as well as a data management portal. Dataset from VICINITY could be used and exploited anonymized from another European project. Dataset from sensors deployed at seniors' houses will provide added value and be the base for other research projects (e.g. statistical data). VICINITY could have an open portal / repository on its website, providing anonymized data's information like timestamp and description.</i></p>
<p>Embargo periods (if any)</p>	<p><i>None</i></p>
<p>Archiving and preservation (including storage and backup)</p>	
<p>Data storage (including backup): where? For how long?</p>	<p><i>Due to ethical and privacy issues, data will be stored in a database scheme at the headquarters of MPH, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by MPH in a secured place. Data will be kept indefinitely allowing statistical analysis.</i></p>

Table 15: Dataset description of the CERTH Motion Sensor

8. Datasets for intelligent mobility from Hafenstrom AS (HITS)



HITS will provide the user requirements specifications and demonstration of transport domain use case, while it will actively participate in the dissemination and exploitation activities of the project. By employing knowhow within standardization bodies, mobility and smart city governance, HITS will allow municipalities and smart cities to better utilize internal resources and improve on services offered to citizens and agencies alike.

Furthermore, HITS will be responsible for the Use cases “Virtual Neighbourhood of Buildings for Assisted Living integrated in a Smart Grid Energy Ecosystem” and “Virtual Neighbourhood of Intelligent (Transport) Parking Space”. Towards this direction, it will be the main partner to bring/arrange the required infrastructure, in collaboration with other Consortium partners (i.e., TINYM partner), for the use case demonstration.

DS.HITS.01.Parkingsensor

Data Identification	
Dataset description	<i>The sensors will be installed at a test site, and will register proximity of objects of a certain size. Future subset may contain information about temperature, humidity, noise, light and other temperature, visual and touch related data. The sensors main task is to detect if the space is occupied. This information will later on be integrated with identification in order to verify that the vehicle/unit that occupies the space is licenced through either booking or ticketing action being taken.</i>
Source (e.g. which device?)	<i>The dataset will be collected through a sensor that is mounted at the parking site.</i>
Partners services and responsibilities	
Partner owner of the device	<i>The device will be the property of the test site owners, where the data collection is going to be performed</i>
Partner in charge of the data collection (if different)	<i>HITS</i>
Partner in charge of the data analysis (if different)	<i>HITS</i>
Partner in charge of the data storage (if different)	<i>HITS</i>
WPs and tasks	<i>The data are going to be collected within activities of WP7 and WP8.</i>
Standards	

Info about metadata (Production and storage dates, places) and documentation?	<i>The dataset will be accompanied with the respective documentation of its contents. Indicative metadata include: (a) description of the experimental setup (e.g. process system, date, etc.) and procedure which is related to the dataset (e.g. proactive maintenance action, unplanned event, nominal operation. etc.), (b) scenario related procedures, state of the monitored activity and involved workers, involved system etc.</i>
Standards, Format, Estimated volume of data	<i>The data will be stored at XML format and are estimated to be 50-300 MB per month.</i>
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Registering parking activity based upon availability, vehicle, ownership/licence, comparing with nearby infrastructure and surrounding ITS technology.</i>
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<i>The full dataset will be available to participants in the project. If the dataset or specific portions of it (e.g. metadata, statistics, etc.) are decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section. Of course these data will be anonymized, so as not to have any potential ethical issues with their publication and dissemination.</i>
Data sharing, re-use and distribution (How?)	<i>The created dataset could be shared by using open APIs through the middleware as well as a data management portal.</i>
Embargo periods (if any)	<i>None</i>
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	<i>Data will be stored in the storage device of the developed system (computer). A back up will be stored in an external storage device. Data will be kept indefinitely allowing statistical analysis.</i>

Table 16: Dataset description of the HITS parking sensor

DS.HITS.02.SmartLight	
Data Identification	
Dataset description	<i>Smart lights will be installed at the lab, and will demonstrate how light and colours can indicate the state of access and availability. Future subset may contain information about proximity, movement, heat sensing (infrared), sound sensing and door contact sensors. The smart lights main task is to visually inform about the state of the parking space. This information may later on be integrated with indicators for occupancy, time to availability and validity.</i>
Source (e.g. which device?)	<i>The dataset will be received from a laptop in the lab.</i>
Partners services and responsibilities	
Partner owner of the device	<i>The device will be the property of the test site owners, where the data collection is going to be performed</i>

Partner in charge of the data collection (if different)	HITS
Partner in charge of the data analysis (if different)	HITS
Partner in charge of the data storage (if different)	HITS
WPs and tasks	<i>The data are going to be collected within activities of WP7 and WP8.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>The dataset will be accompanied with the respective documentation of its contents. Indicative metadata include: (a) description of the experimental setup (e.g. process system, date, etc.) and procedure which is related to the dataset (e.g. proactive maintenance action, unplanned event, nominal operation. etc.), (b) scenario related procedures, state of the monitored activity and involved workers, involved system etc.</i>
Standards, Format, Estimated volume of data	<i>The data will be stored at XML format and are estimated to be 50-300 MB per month.</i>
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Registering parking activity based upon availability, vehicle, ownership/licence, comparing with nearby infrastructure and surrounding ITS technology.</i>
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<i>The full dataset will be available to the members of the consortium. If the dataset or specific portions of it (e.g. metadata, statistics, etc.) are decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section. Of course these data will be anonymized, so as not to have any potential ethical issues with their publication and dissemination.</i>
Data sharing, re-use and distribution (How?)	<i>The created dataset could be shared by using open APIs through the middleware as well as a data management portal.</i>
Embargo periods (if any)	<i>None</i>
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	<i>Data will be stored in the storage device of the developed system (computer). A back up will be stored in an external storage device. Data will be kept indefinitely allowing statistical analysis.</i>

Table 17: Dataset description of the HITS Smart lighting

DS.HITS.03.LaptopTeststation

Data Identification

Dataset description	<i>The laptop test station will be installed at the workbench where the operator normally works, and will aggregate data and process information received wirelessly from other devices delivering data of relevance to the mobility domain and parking in particular. Future subset may contain information about other domains – energy, and data packages from smart home and health-devices. The test stations main task is to process data and trigger activate and log actions accordingly.</i>
Source (e.g. which device?)	<i>The dataset will be collected wirelessly and via USB ports.</i>
Partners services and responsibilities	
Partner owner of the device	<i>The device will be the property of the test site owners, where the data collection is going to be performed</i>
Partner in charge of the data collection (if different)	<i>HITS</i>
Partner in charge of the data analysis (if different)	<i>HITS</i>
Partner in charge of the data storage (if different)	<i>HITS</i>
WPs and tasks	<i>The data are going to be collected within activities of WP7 and WP8.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>The dataset will be accompanied with the respective documentation of its contents. Indicative metadata include: (a) description of the experimental setup (e.g. process system, date, etc.) and procedure which is related to the dataset (e.g. proactive maintenance action, unplanned event, nominal operation. etc.), (b) scenario related procedures, state of the monitored activity and involved workers, involved system etc.</i>
Standards, Format, Estimated volume of data	<i>The data will be stored at XML format and are estimated to be 50-300 MB per month.</i>
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Registering parking activity based upon availability, vehicle, ownership/licence, comparing with nearby infrastructure and surrounding ITS technology.</i>
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<i>The full dataset will be available to the members of the consortium. If the dataset or specific portions of it (e.g. metadata, statistics, etc.) are decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section. Of course these data will be anonymized, so as not to have any potential ethical issues with their publication and dissemination.</i>
Data sharing, re-use and distribution (How?)	<i>The created dataset could be shared by using open APIs through the middleware as well as a data management portal.</i>
Embargo periods (if any)	<i>None</i>
Archiving and preservation (including storage and backup)	

Data storage (including backup): where? For how long?	<i>Data will be stored in the storage device of the developed system (computer). A back up will be stored in an external storage device. Data will be kept indefinitely allowing statistical analysis.</i>
---	--

Table 18: Dataset description of the HITS laptop test station

DS.HITS.04.Sensio_sensors_temperature_motion_lock	
Data Identification	
Dataset description	<i>Sensors for measuring temperature, motion detection and identifying status of door/window lock will be installed in apartments that are managed by caretakers employed by Tromsø municipality. The datasets will contain general information about activities, and offer insight that building manager, caretakers and medical staff can utilize to offer better service and trigger messages should deviations situations occur.</i>
Source (e.g. which device?)	<i>The dataset will be received from a Sensio gateway that stores the data on an external server, and made available to a laptop at the pilot site through an API.</i>
Partners services and responsibilities	
Partner owner of the device	<i>The device will be the property of the test site owners, where the data collection is going to be performed</i>
Partner in charge of the data collection (if different)	<i>HITS</i>
Partner in charge of the data analysis (if different)	<i>HITS</i>
Partner in charge of the data storage (if different)	<i>HITS</i>
WPs and tasks	<i>The data are going to be collected within activities of WP7 and WP8.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>The dataset will contain information on location, and be accompanied with the respective documentation of its contents. Indicative metadata include: scenario related procedures, state of the monitored activity and involved workers, involved system etc.</i>
Standards, Format, Estimated volume of data	<i>The data will be stored at XML format and are estimated to be 30-50 MB per month.</i>
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Identifying usage history used for resource planning and detecting unexpected activities based on activity or lack of activity, as well as measured values versus expected data.</i>

Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<i>The full dataset will be available to the members of the consortium. Specific portions will be accessible to building managers and medical staff. Parts of the data will be anonymised, while other will available through a two-pass data management porta. For privacy reasons, the data access will be limited, so configuration will be made in close cooperation with the service provider.</i>
Data sharing, re-use and distribution (How?)	<i>Due to confidentiality, the created dataset will only be made accessible through a data management portal that is open to medical staff and managers.</i>
Embargo periods (if any)	<i>None</i>
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	<i>Data will be stored in the storage device of the developed system (computer). A back up will be stored in an external storage device. Data will be kept indefinitely allowing statistical analysis.</i>

Table 19: Dataset description of the Sensio sensors

DS.HITS.05.Gorenje_Smart_Appliances_Sensor

Data Identification	
Dataset description	<i>The Gorenje smart appliances installed at the Tromsø pilot site includes a fridge and an oven. The appliances are managed by caretakers employed by Tromsø municipality, the tenants themselves and the building manager. The appliances contain sensors that among other things can measure timestamps and temperature.</i> <i>The data harvested will be used to identify usage history in order to offer better service, identify abnormal behaviour, and otherwise generate logs that can be used for statistical analysis.</i>
Source (e.g. which device?)	<i>The data will be collected by specific smart appliances (i.e. oven, fridge) provided by Gorenje and adjusted to VICINITY requirements. The data will be made available to a laptop at the pilot site through an API.</i>
Partners services and responsibilities	
Partner owner of the device	<i>The device will be the property of the test site owners, where the data collection is going to be performed.</i>
Partner in charge of the data collection (if different)	<i>HITS, GORENJE</i>
Partner in charge of the data analysis (if different)	<i>HITS, GORENJE</i>
Partner in charge of the data storage (if different)	<i>HITS</i>
WPs and tasks	<i>The data are going to be collected within activities of WP6, WP7 and WP8.</i>
Standards	

Info about metadata (Production and storage dates, places) and documentation?	<i>The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.</i>
Standards, Format, Estimated volume of data	<i>A collection of data is dispatched every 15 minute. The format is based on standards provided by Gorenje.</i>
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Usage data to identify behaviour patterns and as mean for training disabled users in being more self-sufficient are examples are examples of value-added services that can be built on top of the platform. As the data pool increases, more services are expected to be included.</i>
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<i>The full dataset of Gorenje devices deployed at the Tromsø pilot site “Teaterkvarteret 1. Akt”, will be stored at the Gorenje Cloud in a local experimental repository. The full dataset will be available to selected members of the consortium. Specific portions will be accessible to building managers and medical staff. Parts of the data will be anonymised, while other will available through a two-pass data management porta. For privacy reasons, the data access will be limited, so configuration will be made in close cooperation with the service provider.</i>
Data sharing, re-use and distribution (How?)	<i>Anonymised parts of the dataset will be available for training and statistic purposes. Aggregated data that could be used to identify the user or other privacy related information will be limited. Due to confidentiality, the created dataset will only be made accessible through a data management portal that is open to medical staff and managers.</i>
Embargo periods (if any)	<i>None</i>
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	<i>Data will be stored in the storage device of the developed system (computer). A back up will be stored in an external storage device. Data will be kept indefinitely allowing statistical analysis.</i>

Table 20: Dataset description of the Gorenje smart appliances sensor

9. Datasets for buildings from Tiny Mesh AS (TINYM)

Tinymesh

The primary role of Tiny Mesh Company is as a developer and technology provider, with the company’s IoT solution as the main enabling technology. The goal is to offer promising technology solutions through participation in use cases. We focus on creating new products, services and business model as part of the Internet-of-Everything (IoE). New potential arise when IoE is used for connecting, integrating and controlling all kinds of meters, street lights, sensors, actuators, assets, devices, tags and other devices.

TINYM will contribute in the practical implementation through their work with definitions of use case. TINYM will take practical ownership of the various demo sites through the role as of leader of WP7.

DS. TinyMesh.01.Door_Sensor

Data Identification	
Dataset description	<i>The sensors will be installed in the door of a room where there is a need for monitoring usage. Data packet contains sensor data of movement.</i>
Source (e.g. which device?)	<i>Discrete digital input</i>
Partners services and responsibilities	
Partner owner of the device	<i>The property owner Tiny-Mesh</i>
Partner in charge of the data collection (if different)	<i>Tiny-Mesh</i>
Partner in charge of the data analysis (if different)	<i>Tiny-Mesh</i>
Partner in charge of the data storage (if different)	<i>Tiny-Mesh</i>
WPs and tasks	<i>The data are going to be collected within activities of WP7 and WP8.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>Metadata about location of the sensor, network topology and network status will be available in Tiny-Mesh Workbench.</i>
Standards, Format, Estimated volume of data	<i>Data is delivered as a discrete value indicating if door has been opened or closed, volume of data depends on the usage.</i>
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>The purpose of this collection is to give input data for analysis of room usage for analyses to the building owner and Facility manager.</i>

Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<i>Data access for building manager and facility manager. Data will be anonymized, so as not to have any potential ethical issues with their publication and dissemination.</i>
Data sharing, re-use and distribution (How?)	<i>Data access is confidential. Only members of the consortium, building manager and facility manager will have access on it for privacy reasons.</i>
Embargo periods (if any)	-
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	<i>Unless specified otherwise by client the data will be stored in a Value Added Service.</i>

Table 21: Dataset description of the Tiny-Mesh Door Sensor

DS. TinyMesh.02.Energy_Water_Consumption_Sensor

Data Identification	
Dataset description	<i>The sensors will be installed to measure consumption of water and electronics.. Data packet contains sensor data of movement.</i>
Source (e.g. which device?)	<i>Data is retrieved through industry-standard meters and communicated through Tiny-Mesh infrastructure before being made available to the consortium.</i>
Partners services and responsibilities	
Partner owner of the device	<i>The property owner Tiny-Mesh</i>
Partner in charge of the data collection (if different)	<i>Tiny-Mesh</i>
Partner in charge of the data analysis (if different)	<i>Tiny-Mesh</i>
Partner in charge of the data storage (if different)	<i>Tiny-Mesh</i>
WPs and tasks	<i>The data are going to be collected within activities of WP7 and WP8.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>Metadata about location of the sensor, network topology and network status will be available in Tiny-Mesh Workbench.</i>
Standards, Format, Estimated volume of data	<i>Communication with meter will be on a proprietary interface according to meter vendor. Data will be delivered as KW/h or l/h on a configurable interval of (default: 1 minute).</i>
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>The purpose of this collection is to give input data for analysis of resource usage to control peak electricity or alarm of abnormal use.</i>

Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<i>Data access is restricted to the consortium, building manager and facility manager. Data will be anonymized, so as not to have any potential ethical issues with their publication and dissemination.</i>
Data sharing, re-use and distribution (How?)	<i>Data access is confidential. Only members of the consortium, building manager and facility manager will have access on it for privacy reasons.</i>
Embargo periods (if any)	-
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	<i>Data will be stored in the metering devices as well as the Tiny-Mesh provided Value Added Service.</i>

Table 22: Dataset description of the Tiny-Mesh consumption sensor for energy and water.

DS. TinyMesh.03 Tinymesh_Gateway

Data Identification	
Dataset description	<i>Data packed from any Tinymesh network</i>
Source (e.g. which device?)	<i>The Tiny-Mesh Gateway relays information from different Tiny-Mesh devices to upstream service.</i>
Partners services and responsibilities	
Partner owner of the device	<i>Tiny-Mesh</i>
Partner in charge of the data collection (if different)	<i>Tiny-Mesh</i>
Partner in charge of the data analysis (if different)	<i>Tiny-Mesh</i>
Partner in charge of the data storage (if different)	<i>Tiny-Mesh</i>
WPs and tasks	<i>The data are going to be collected within activities of WP7 and WP8.</i>
Standards	
Info about metadata (Production and storage dates, places) and documentation?	<i>Tiny-Mesh Gateway is a serial communication device that can transfer data in two modus; transparent and packed.</i>
Standards, Format, Estimated volume of data	-
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	<i>Tiny-Mesh Gateway is serial communication device.</i>

Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	<i>The full dataset will be confidential and only the members of the consortium will have access on it.</i>
Data sharing, re-use and distribution (How?)	<i>Data and metadata will be accessible by an API in Tinymesh Cloud.</i>
Embargo periods (if any)	-
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	<i>Data and metadata will be accessible by an API in Tinymesh Cloud.</i>

Table 23: Dataset description of the Tiny-Mesh gateway

10. Conclusions

This document is the second version of the Data Management Plan. It is based on knowledge harvested through describing requirements, preparing the VICINITY architecture, and planning the pilot sites. The updated datasets have been delivered from the participants that are responsible for the test labs and the living labs, and describes procedures and infrastructure that have been defined at this point in the project.

The work on semantics and privacy issues has continued. It is the process of clarification of procedures that has led to many of the updates that are found in this document. Certain areas still need some attention. This will in particular matter for Open Calls, as these are still tentative and the documents and other material is still being worked out by the VICINITY consortium. Activities for a Data Management Portal have proceeded, and a demonstration has been held twice that presented how the VICINITY architecture works, how it integrates and how the concept of virtual neighborhood functions in practical terms. More updates is envisaged after studies of the pilot sites proceeds, and open calls are being presented. Future versions may have updated Consent forms as well since the upcoming GDPR may lead to changes in how privacy and ethics issues are formulated.

Lessons learned from this report is there has been introduced more IoT assets that will be integrated within the ecosystems that will be tested. There has been a fruitful discussion between project partners, which increases the quality of this document. Ownership of data become more important, and will receive special attention in the next part. The Data Management Portal is still under work, but need for each project partner to contribute to editing / access rights will need to be managed accordingly. It must also be noted that the partners are unable to exactly specify what kind of datasets that will be relevant as the project proceeds. This is what they expect to learn from the pilot sites and other tests conducted at the workbench. It is therefore expected that the datasets may change accordingly.

The VICINITY Data Management Plan still put a strong emphasis of the appropriate collection – and publication should the data be published – of metadata, storing all the information necessary for the optimal use and reuse of those datasets. This metadata will be managed by each data producer, and will be integrated in the Data Management Portal. This is considered even more important with the upcoming deployment of the General Data Privacy Regulations (GDRP).

The final version of DMP is due in December 2019. It is expected to present the final datasets and lessons learned, alongside plans for further management of test data and production data. It will provide information on the existence (or not) of similar data and the possibilities for integration and reuse. In addition, issues like the period of data preservation, the approximated end volume, the associated costs and how these are planned to be covered will be tackled in order to make the Portal and other necessary management tools operational and to provide a detailed Management Plan for each dataset.

References

- European Commission. (2013). Guidelines on Data Management in Horizon 2020. Retrieved 2 June, 2015, from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- European IPR Helpdesk. (2014). Fact Sheet Open Access to publications and data in Horizon 2020: Frequently Asked Questions (FAQ). Retrieved 3 July, 2015, from https://www.iprhelpdesk.eu/sites/default/files/newsdocuments/Open_Access_in_H2020.pdf
- H2020 Programme: Guidance - How to complete your ethics self-assessment http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-self-assess_en.pdf
- Roles and Functions of Ethics Advisors/Ethics Advisory Boards in EC-funded Projects http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/ethics-guide-advisors_en.pdf
- GDPR: Testdata på ville veier (Norwegian newsarticle) <https://www.digi.no/artikler/kronikk-gdpr-testdata-pa-ville-veier/410127>
- Hva betyr GDPR for oss som lager IT-løsninger? (Norwegian newsarticle) <https://www.digi.no/artikler/kronikk-hva-betyr-gdpr-for-oss-som-lager-it-losninger/410419>
- Let's Cut The Crap On GDPR by Carl Gottlieb <https://www.youtube.com/watch?v=A4kfSxKdJVE>
- FmhDND: Lansering av Datatilsynets veileder for innebygd personvern (GDPR) <https://www.youtube.com/watch?v=e7WmVSaFTfc>
- VICINITY project at "Boring but lucrative, the real Internet of Things" <http://vicinity2020.eu/vicinity/content/vicinity-project-%E2%80%9Cboring-lucrative-real-internet-things%E2%80%9D-cambridge-wireless>
- <http://www.eugdpr.org/> and <http://www.eugdpr.org/gdpr-faqs.html>
- Checklist of data protection best practices <https://blogs.thomsonreuters.com/answerson/data-protection-action-items-for-firms/>
- Security & Privacy Best Practices <https://otalliance.org/resources/security-privacy-best-practices>
-

Annex 1 – Preferred Formats

3.13. Selection of File Formats

All formats of digital files stand the risk of becoming obsolete in the future. If a file format becomes obsolete, it means that the current software will not be able to represent and use the content of the file in the way it was meant to at the time of creation.

To prevent file format obsolescence, some precautions can be taken. One such measure is to select file formats which have a high chance of remaining usable in the far future. As a general guideline, VICINITY considers that the file formats best suited for long-time preservation and accessibility:

1. are commonly used;
2. have open specifications;
3. are independent of specific software, developers or suppliers.

However, it is not always possible to select formats that meet with all of these ideal attributes.

3.14. Preferred and Acceptable Formats

VICINIY has assessed a number of file formats resulting in a list of preferred formats and acceptable formats. This list will change over time as new formats will be developed and others will fall into disuse.

Preferred formats are the file formats which can be trusted to offer the best long-term guarantees for usability, accessibility and robustness. In principle, VICINITY expects these formats to be durable for the long term. VICINITY will accept research data deposited in preferred formats in VICINITYs repository without question.

Acceptable formats are file formats which are commonly used besides the preferred formats; have average to reasonable scores regarding their usability, accessibility and robustness in the long term. VICINITY strongly prefers the use of preferred formats but in most cases, the use of acceptable formats will be allowed in to the archive as well.

	Preferred format(s)	Acceptable format(s)
Text documents	PDF/A (.pdf)	OpenDocument Text (.odt) MS Word (.doc, .docx) Rich Text File (.rtf) PDF (.pdf)
Text file	Unicode TXT (.txt, ...)	Non-Unicode TXT (.txt, ...)
Marked-up language		XML (.xml) HTML (.html)
Spreadsheets	OpenDocument Spreadsheet (.ods) Comma Separated Values (.csv)	MS Excel (.xls, .xlsx) PDF/A (.pdf) OOXML (.docx, .docm)
Databases	ANSI SQL (.sql, ...) Comma Separated Values (.csv)	MS Access (.mdb, .accdb) dBase III or IV (.dbf)

Statistical data	R SPSS Portable (.por) SAS transport (.sas) STATA (.dta)	
Images (raster)	JPEG (.jpg, .jpeg) TIFF (.tif, .tiff) PNG (.png)	JPEG 2000 (.jp2)
Images (vector)	Scalable Vector Graphics (.svg)	Adobe Illustrator (.ai) PostScript (.eps)
Video	MPEG-2 (.mpg, .mpeg, ...) MPEG-4 H264 (.mp4) Lossless AVI (.avi) QuickTime (.mov)	
Audio	WAVE (.wav)	MP3 AAC (.mp3)
Computer Design (CAD)	Aided AutoCAD DXF versie R12 (.dxf)	AutoCAD other versions (.dwg, .dxf)
Geographic Information (GIS)	Geographic Markup Language (.gml) MapInfo Interchange Fomat (.mif/.mid)	ESRI Shapefiles (.shp and associated files) MapInfo (.tab and associated files) Keyhole Markup Language (.kml)
Images (georeferenced)	GeoTIFF (.tif, .tiff)	TIFF World File (.tfw en .tif)
Raster GIS	ASCII GRID (.asc, .txt)	ESRI GRID (.grd and associated files)
3D	WaveFront Object (.obj) X3D (.x3d) STEP (standard filetype for CNC machining)	COLLADA (.dae) Autodesk FBX (.fbx) STL (standard filetype for 3D printing)
RDF	W3C standards	

Table 24: summarised overview of VICINITYs preferred and acceptable Formats

Annex 2: VICINITY Consent form Template

A template of the consent form to be used is presented below, to be adopted as required per pilot use case.



Consent Form

Purpose of the study
A commonly understandable written description of the project and its goals (2-3 paragraphs)

Planned Project Progress
The planned project progress and the related testing and evaluation procedures (1-2 paragraphs)

Disclaimer Rights
Advice on unrestricted disclaimer rights on their agreement.

Voluntary Participation Form for the needs of the VICINITY project

1. Participant Information
Basic information and participant's reference code ID (the reference code ID will be used throughout the pilot trial execution)

2. Study Information
Details about the pilot Use Case

3. Participant's Questionnaire
Questions verifying that the participant:

- *has been fully informed on the purpose, duration, procedures of the study;*
- *has been informed on the rights to deny participating or to quit from the study and about the corresponding consequences.*
- *has been informed on the contact person in case that I have questions and queries about the study.*
- *had adequate time to make my decision concerning my participation in the study.*
- *comprehend that he/she can quit from the study at any time without having to justify his/her decision.*
- *has been informed about potential effects, difficulties and dangers.*
- *has been informed about the sensors equipment that will be used to collect data.*
- *has been informed about the security of the study data and results.*
- *has been ensured about the confidentiality of his/her personal information. Publications of the study results do not allow the personal data recognition, due to the principle of anonymity. Always under the confidentiality principles.*

4. Signed Consent to Participate
A signed consent of the participant allowing the study responsible to examine and inspect the data collected during the study.

Annex 3 – The Ethical Advisory Board

Ethics is an integral part of the VICINITY project. Although the VICINITY model does not expose, use or analyze personal sensitive data, the consortium is aware that a number of privacy and data protection issues could be raised by activities within the scope of the project. The need for clear guidelines and support from the organization is important.

VICINITY consortium respects the ethical rules and standards of H2020, and those reflected in the Charter of Fundamental Rights of the European Union. VICINITY will address any ethical issues in WP10 (T10.2) as well as allocated specific task 6.4 to review the deployed solutions for privacy and security. Thus VICINITY will assure the investigation, management and monitoring of ethical and privacy issues that could be relevant to its envisaged technological solution and will establish a close-cooperation with the Ethics Helpdesk of the European Commission.

The European Commission is requesting that projects are “ethics-ready”. This will be done in full compliance with any European and national legislation and directives relevant to the country where the data collections are taking place (INTERNATIONAL/EUROPEAN):

- The Universal Declaration of Human Rights and the Convention 108 for the Protection of Individuals with Regard to Automatic Processing of Personal Data and
- Directive 95/46/EC & Directive 2002/58/EC of the European parliament regarding issues with privacy and protection of personal data and the free movement of such data.

For this purpose, the VICINITY consortium has established an ethical advisory board (EAB). The members of EAB consists of partners from a wide area of expertise.

As of November 2017, the EAB consist of:

Partner (Short)	Name	Surname	Roles
UNIKL	Christoph	Grimm	
UNIKL	Carna	Radojicic	
CERTH	Thanasis	Tryferidis	
CAL	Nigel	Wall	
TINYM	Erik	Nygaard	
ENERC	Natalie	Samovich	Manager of board
MPH	Alexandra	Ananika	

Table 25: VICINITY ethical advisory board (EAB)

Annex 4 – Contacts technical data and internal training

The partners that are responsible for the living labs have assigned a contact person to handle description of datasets and that serves as a contact point for technical information. Additionally, this person will be responsible for ensuring internal knowledge about the technical aspects of the pilot site installations and integrations.

Partner (Short)	Name	Surname	Contact
AAU	Yajuan	Guan	ygu@et.aau.dk
ENERC	Marc	Rechter	m.rechter@enercoutim.eu
GNOMON	Kostis	Kaggelides	k.kaggelides@gnomon.com.gr
HITS	Flemming	Sveen	flsveen@online.no
TINYM	Mariann	Sundvor	mariann@tiny-mesh.com

Table 26: Partners contacts for technical information and internal training

Annex 5 – Assessment tools

These tables are based on recommendations described in a paper ⁵by the European Commission to ensure that the DMP conforms to EU requirements (Article 69). Not all elements apply to VICINITY as a project, but do still present a sound framework for QA activities to improve on the document.

This table serve as a checklist for the annual review and adjustment of the DMP.

	Status	Score
<p>General principles and ethics and security</p> <p><i>Comment:</i> The general principles are well defined. There is expected some updates will take place at the end of the test pilots. This also applies to ethics and security issues, but will be for a large degree be based on feedback from pilot site stakeholders.</p>	√	4/5
<p>IPR management and security, production data and test data</p> <p><i>Comment:</i> The IPR management section covers all of the issues that so far has been identified. Some of these topics also address how data is generated and further use. Proper guidelines are laid out, and has been accepted by all parties involved.</p>	√	5/5
<p>Personal data protection and Data Management Portal</p> <p><i>Comment:</i> All concerns are addressed, and GDPR has laid the foundation that data exchange – ownership and visibility adheres to.</p>	√	4/5
<p>Open Calls and data sharing</p> <p><i>Comment:</i> This is a new section that is still being worked out by the project partners. Some of the information presented here is still somewhat premature, and will most likely be updated when the Open Calls have been launched.</p>	√	3/5
<p>Standards, metadata and archiving</p> <p><i>Comment:</i> Description of semantic data and storage lay out routines and responsible partners in a structured way.</p>	√	5/5
<p>Datase</p> <p><i>Comment:</i></p>	√	5/5

⁵ http://ec.europa.eu/regional_policy/archive/country/commu/2000-2006/docoutils/fiches/EN/03_EN.pdf

The project participants have all delivered datasets that are well defined and described, and serve as a good starting point for further development of the Data Management Portal and configuration on the pilot sites.



List of Tables

Table 1: Best practice for use of production data	16
Table 2: datasets stored in the VICINITY management portal	18
Table 3: Format of dataset description	19
Table 4: Dataset description of the AAU GRID status	25
Table 5: Dataset description of the ENERC METEO station	27
Table 6: Dataset description of the ENERC building status	28
Table 7: Dataset description of the ENERC grid status	30
Table 8: Dataset description of the GNOMON pressure sensor	32
Table 9: Dataset description of the GNOMON weight sensor	34
Table 10: Dataset description of the GNOMON fall sensor	36
Table 11: Dataset description of the GNOMON Wearable Fitness Tracker Sensor	38
Table 12: Dataset description of the GNOMON Beacon Sensor	39
Table 13: Dataset description of the GNOMON/CERTH Gorenje Smart Appliances Sensor	41
Table 14: Dataset description of the CERTH Occupancy Sensor	43
Table 15: Dataset description of the CERTH Motion Sensor	45
Table 16: Dataset description of the HITS parking sensor	47
Table 17: Dataset description of the HITS Smart lighting	48
Table 18: Dataset description of the HITS laptop test station	50
Table 19: Dataset description of the Sensio sensors	51
Table 20: Dataset description of the Gorenje smart appliances sensor	52
Table 21: Dataset description of the Tiny-Mesh Door Sensor	54
Table 22: Dataset description of the Tiny-Mesh consumption sensor for energy and water	55
Table 23: Dataset description of the Tiny-Mesh gateway	56
Table 24: summarised overview of VICINITY's preferred and acceptable Formats	60
Table 25: VICINITY ethical advisory board (EAB)	62
Table 26: Partners contacts for technical information and internal training	63

List of Figures

Figure 1: Data Management Plan – deliverables 2016 – 2019	9
Figure 2: Domains and some of the functionalities the DMP has to cover	10
Figure 3: Example of potential data points in use cases that generate data	11
Figure 4: The VICINITY consortium includes partners from different sectors with confidential data	12
Figure 5: VICINITY complies with European and national legislations	13
Figure 6: The VICINITY architecture is decentralised by design	15
Figure 7: Involving 3rd parties through open calls will provide VICINITY with valuable experience, and evolve interoperability	20
Figure 8: Datasets will be prepared and provided in XML and JSON format	20
Figure 9: Example of SysML model of Virtual Oslo Science City	22
Figure 10: Data will only be provided partners with proper login credentials	22