



Grant Agreement No.: 680708

Project acronym: HIT2GAP

Project title: Highly Innovative building control Tools Tackling the energy performance gap

Research and Innovation Action

Topic: EeB-07-2015 New tools and methodologies to reduce the gap between predicted and actual energy performances at the level of buildings and blocks of buildings

Starting date of project: 1st of September 2015

Duration: 48 months

D1.8 – Data Management Plan

Organisation name of lead contractor for this deliverable: NOBATEK		
Version 1 – Rev.0	Due Date	M6 - 29 th of February 2016
	Submission Date	29 th of February 2016
	Primary Authors	NOBATEK - P. BRASSIER, G. TEKLI
	Contributors	All partners
	Reviewers	Nikos SAKKAS (API), Richard CHBEIR (LIU)

Revision table for this deliverable:		
Version 2 – Rev.0	Submission Date	07/04/2017
	Revisor	P. BRASSIER (NBK), P. BOURREAU (NBK), L.M. Blanes Restoy (NUIG)

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

Table of Contents

Executive Summary	4
1 Introduction	5
1.1 Purpose	5
1.2 Contributions of partners.....	6
1.3 Baseline definitions and Concepts	7
2 Policy on scientific publications	8
3 Data management methodology used in the HIT2GAP project	10
3.1 DMP Management structure	10
3.2 Data management process	12
3.3 Categories of datasets	12
3.4 Repository options and selection	13
3.5 Lifecycle of the Data management.....	14
3.5.1 Dataset management template	14
3.5.2 Generalities on data workflow in the HIT2GAP platform.....	16
4 Data set reference, name and description	18
4.1 Collected data.....	18
4.2 Generated data.....	19
4.3 Dataset management template	19
4.4 Datasets identified within HIT2GAP	21
4.4.1 Building-related: measurements	21
4.4.2 Building-related: description.....	23
4.4.3 Users-related: measurements	25
4.4.4 Users-related: preferences.....	26
4.4.5 Generated by modules.....	27
4.4.6 Internal: needs and users requirements	29
4.4.7 Internal: description of partners' software in HIT2GAP project	30
4.4.8 Internal: Market exploitation data	31
5 Conclusions	32
6 References	33
7 Glossary and acronyms	34
8 Annex: Guidelines related with the process for open access data within HIT2GAP	35
8.1 H2020 Open Access policy	35
8.2 Open Access Options	35
8.3 Open access repositories	36
8.4 Open source licensing options	37
8.5 Resources	38

List of Tables

TABLE 1: SCIENTIFIC PUBLICATIONS, COMMUNICATIONS, REPORTS AND RELATED DATASETS GENERATED IN THE FRAME OF THE HIT2GAP PROJECT	9
TABLE 2: DMP MANAGEMENT STRUCTURE.....	10
TABLE 3: H2020 DATASET TEMPLATE	15
TABLE 4: TECHNICAL CLASSIFICATION OF COLLECTED AND GENERATED DATA	16
TABLE 5: DATA COLLECTED IN THE FRAME OF THE HIT2GAP PROJECT	18
TABLE 6: TECHNICAL DATA GENERATED THROUGH THE DEVELOPMENT AND USE OF THE TOOLS DEVELOPED WITHIN HIT2GAP	19
TABLE 7: DATA MANAGEMENT TEMPLATE USED IN HIT2GAP	20

Executive Summary

In accordance with H2020 guidelines related to data management [1], every participant projects to the **pilot action on open access to research data** should develop a Data Management Plan (DMP) in which they specify what data will be open. According to these guidelines, a DMP should detail:

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

In that frame, this document constitutes the Data Management Plan and strategy for the open data treatment within the HIT2GAP project. The document also specifies the data that are maintained confidential and explains why. It should be noted that this document belongs to the Work Package number 1 of the HIT2GAP project which is responsible for collecting the requirements. The overall requirements of the project are directly related to the collected data in the project that will be used by data processing modules for energy saving purposes. These modules will generate new data related to the energy saving and can feed energy management tools developed during the project lifespan.

This document is likely to evolve during the project lifecycle and be updated to include new data sets and new results sets.

This version of the document is the second version corresponding to the update expected at the end of the first period of the project (M18). It includes the refined list of data collected and generated as well as solutions for the repository of data sets selected within the HIT2GAP project.

1 Introduction

1.1 Purpose

In Horizon 2020, a limited and flexible pilot action on open access to research data is currently implemented. In accordance with H2020 guidelines related to data management [1], every participant to the **pilot action on open access to research data** should develop a Data Management Plan (DMP) in which he/she specifies:

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

At the proposal stage, the HIT2GAP consortium decided to participate in the Pilot on Open Research Data in Horizon 2020 on a voluntary basis. In that frame, the deliverable D1.8 provides the main requirements concerning the open access data and explains how this new requirement is addressed in the HIT2GAP project.

As specified in the guidelines, a DMP describes the data management life cycle for all data sets that will be collected, processed or generated by the research project. It is a document outlining how research data will be handled during a research project, and even after the project is completed, describing what data will be collected, processed or generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved.

It should be noted that this document is also closely related to the ethical issues management plan that also requires some technical restrictions associated to the security of personal data that will be collected in the pilot sites. Moreover, this DMP is consistent with the exploitation and protection of results that will be developed as part of the HIT2GAP project.

In order to elaborate the DMP, the checklist proposed by the Digital Curation Centre (DCC¹) in the UK has been considered, with key questions as follows:

1. What data will you collect or create?
2. How will the data be collected or created?
3. What documentation and metadata will accompany the data?
4. How will you manage any ethical issues?
5. How will you manage copyright and Intellectual Property Rights (IPR) issues?
6. How will the data be stored and backed up during the research?
7. How will you manage access and security?

¹ http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf

8. Which data should be retained, shared, and/or preserved?
9. What is the long-term preservation plan for the dataset?
10. How will you share the data?
11. Are any restrictions on data sharing required?
12. Who will be responsible for data management?
13. What resources will you require to deliver your plan?

The preparation of the present deliverable has consisted in answering all these questions.

The remainder of the document D1.8 is organised as follows:

- **Chapter 1:** Introduction
- **Chapter 2:** Policy on scientific publications
- **Chapter 3:** Data management methodology used in the HIT2GAP project
- **Chapter 4:** Data set reference, name and description. The data sets tables that will be used within the project and that will be updated across the project lifespan.
- **Chapter 5:** Conclusions

Data Management Plan activities are conducted as part of WP1 and are also related to the task 1.2 related to the specifications of the field data infrastructure. The WP1 leader is APINTECH while the T1.2 leader is LIUPPA and they both have been involved in the review of the first version of D1.8. The DMP is also closely related to the pilot sites where the data will be collected and post processed using the tools developed as part of the project. The partners involved in these activities are the one who will collect personal data, and the pilot site managers as well as partners involved in the technical solutions associated to the technical data collection. Partners involved in the modules development and integration are also involved in the DMP as end-users of these collected data and as providers/generators of new data. The Task 7.3 (Scientific diffusion) leader, which is NUIG, is also involved in the DMP focusing on publications. The DMP is also consistent with and builds upon Clause 29 of the Grant Agreement related to Open Access.

1.2 Contributions of partners

NOBATEK has structured the work in this report, made a literature survey in the domain of Data Management and coordinated the collection of the input from the other partners involved in the data management plan within the HIT2GAP project. Pilot sites managers and developers of the modules related with post-processing of data have been involved in the process as well in order to collect their point of view and agreement or not on the openness of the data collected or generated. NUIG has set up the public repository used for the publications and data generated within the project.

1.3 Baseline definitions and Concepts

The baseline document for **data management** in H2020 is “Guidelines on FAIR Data Management in Horizon 2020, Version 3.0, 26 July 2016” [1]. As stated in this document, every participant to the **pilot action on open access to research data** should develop a Data Management Plan (DMP) in which they specify what data will be open, how it will be treated and managed. According to these guidelines, a DMP should be developed and detail:

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

The baseline document for **Open Access** in H2020 is “Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 [2].” According to this document, Open access is defined as follows:

“Open access (OA) can be defined as the practice of providing on-line access to scientific information that is free of charge to the end-user and that is re-usable. ‘Scientific’ refers to all academic disciplines; in the context of research and innovation, ‘scientific information’ can refer to (i) peer-reviewed scientific research articles (published in scholarly journals) or (ii) research data (data underlying publications, curated data and/or raw data).

These definitions describe ‘access’ in the context of open access as including not only basic elements such as the right to read, download and print, but also the right to copy, distribute, search, link, crawl, and mine.”

In relation to the HIT2GAP project, these concepts are applied in the following way:

- 1) HIT2GAP is mandated to provide an Open Access to its scientific publications: this will be achieved by publishing research results in open access mode, and making them available by uploading and indexing them to open access repositories (repositories indexed in the aggregator OpenAire (www.openaire.eu). Additionally, links will be provided in the HIT2GAP web page and partners will be encouraged to disseminate results through more popular web-based platforms (e.g.: www.researchgate.net or www.academia.org);
- 2) HIT2GAP will define a data management plan compliant with the need to provide an open access to some of the other data collected and generated in the project.

2 Policy on scientific publications

With respect to publications, the following options are available:

- Publications are published in **open access** journals. This option is encouraged to all project partners. This means the journal grants open access by definition. These are normally:
 - Conference proceedings (IBPSA, CIBSE Technical Symposium, etc.);
 - Open access journals like ITCon (<http://www.itcon.org/about>) with no Article Processing Charges (APC);
 - Open access journals like Cogent Engineering with "low cost" APCs (<https://www.cogentoa.com/journal/engineering>) by Taylor&Francis.
- Publications are published in **gold open access mode**, that means a higher fee is paid and the publication is granted open access immediately;
- Publications are published in **green open access mode**, that means the publication can be made available after an embargo period, which is set up by the publisher. The publication fee is lower than the gold mode. The embargo period is substantially set for a longer period (18 months, 24 months etc.) than what is required under H2020 (6 months), so this option is not likely to be possible for H2020 so for the journals targeted by the project partners;
- Publications are **not published in open access mode**. This option is not encouraged.

In relation to the HIT2GAP project, these concepts are applied in the following way:

- HIT2GAP has made mandatory for all partners to provide Open Access to their scientific publications and data: this will be done by granting open access to publications and data and uploading them to an open access repository;
- For HIT2GAP participants, the default open repository to upload scientific dissemination publications and datasets is ZENODO (<https://zenodo.org>);
- HIT2GAP is a "community" within ZENODO where all HIT2GAP papers will be linked (<https://zenodo.org/communities/hit2gap/about/t>). Publications and data uploaded in ZENODO are indexed automatically in the OpenAire aggregator;
- Additionally, project partners might want to use other platforms to upload data or publications. Repositories for data are available within the "re3data" registry (www.re3data.org) and OpenAire (www.openaire.eu). Project partners are encouraged to disseminate outputs also in web-based scientific social networks such as www.academia.org or www.researchgate.net ;
- Guidelines have been specified to all participants in order to meet all open access obligations, to meet the Data Management Plan, and to avoid conflict of interest with other participants, as required by Art. 29 of the EU Grant Agreement;

HIT2GAP project participants will pursue default open access publications and GOLD open access publication option in line with EU grant agreement policies. In cases where GOLD open access is not possible, participants will share preprints and/or accepted manuscripts according with the journal policies.

Within the peer reviewed process related to deliverables submission, the peer reviewers will consider the DMP requirements in the review of the documents and will outline specific datasets that should be highlighted and treated in accordance with the methodology proposed in the present document.

Internal guidelines have been elaborated to give all project beneficiaries the description of the process retained for the data open access within the HIT2GAP project (see annex, section 8).

Table 1 provides the list of data identified as scientific publications in the HIT2GAP project, and the associated accessibility strategy.

Table 1: Scientific publications, communications, reports and related datasets generated in the frame of the HIT2GAP project

Type/Category	Initial Identification	Accessibility /Sharing Strategy
Peer reviewed publications	Journal and conference articles related to the developments conducted during the project and the results achieved thanks to these developments.	<ul style="list-style-type: none"> Open Access
Non-Peer reviewed publications	Conference papers; Project presentations at different events (seminars both at national and European levels, workshops, stakeholder's meetings...); Technical journal contributions	<ul style="list-style-type: none"> Open Access
Data used in the publication of scientific productions	Data inputs and outputs for data processing modules in publications; Data stored in the core platform.	<ul style="list-style-type: none"> Open Access
Public information and communication materials related to the HIT2GAP project	Newsletters, leaflets, Brochures, Posters, Press Releases; Public Deliverables.	<ul style="list-style-type: none"> Open Access
Non-public documents related to the project	Confidential project deliverables.	<ul style="list-style-type: none"> Confidential to consortium partners

3 Data management methodology used in the HIT2GAP project

This chapter details the way the project management activities and project management structure are related to the data management activities. It also proposes a data categorisation for the data collected and generated through the project as well as the repository options that will be selected per each category.

3.1 DMP Management structure

As the HIT2GAP project is focusing on added value delivered through advanced treatment of the data collected in a building, data management is underlying all the activities within the project and is therefore reflected in all the project management activities.

The following table summarizes the main contributors in DMP within the project and their role in DMP.

Table 2: DMP Management Structure

Member of the HIT2GAP consortium	Main role in the project	Role in Data Management Activities
NBK	Coordinator WP4 leader WP8 leader	<ul style="list-style-type: none"> • Manages EMDESK platform; • Drafts GM, PMT and GA meeting agendas; • Appoints quality reviewers; • Overall responsible for project management; • Manipulates data in the frame of the platform developed as part of the project; • Ensures privacy and integrity of the data collected during the project; • Develops and manages data management plan; • Ensures that data management is part of project activities including dissemination, deliverables quality review and project meetings; • Coordinates with the dissemination management of the project (BRE); • Coordinates with the scientific diffusion manager (NUIG); • Conducts the ethics issues management of the project.

Member of the HIT2GAP consortium	Main role in the project	Role in Data Management Activities
MOS, BES, GIR, CYL, WRS	Pilot site managers MOS=Pilot manager, WP5 leader and manager of the Polish pilot site BES=manager of the French pilot site GIR=manager of the Spanish pilot site CYL=manager of the Irish pilot site	<ul style="list-style-type: none"> • Manages the link with the pilot sites and aggregates the information related to the pilot sites (in this frame data are related to building design, data measured in the building during operation, data from end-users, data from building owners); • Ensures that the occupants of the pilot sites are aware of the HIT2GAP project objectives and agree in participating to the demonstration phase of the project; • Are responsible for getting the ethical authorizations from the national authorities in charge of data protection in the countries of each pilot site; • Publish data used by modules developers in an open access platform.
R2M	Innovation manager WP6 leader	<ul style="list-style-type: none"> • Identifies and manages the exploitable results (with the support of consortium partners); • Coordinates with the dissemination manager of the project (BRE).
BRE	Communication manager WP7 leader	<ul style="list-style-type: none"> • Defines and implements dissemination strategy and dissemination plan; • Gathers the information related to communications conducted by all the partners; • Uploads public deliverables on the website and as such oversees Open Access Publications and Data; • Collaborates with the activities dedicated to innovation and exploitation (coordinates with R2M);
NUIG	T7.3 leader	<ul style="list-style-type: none"> • Manages publications and related data in the consortium: verifies satisfaction of the publication policy specified in the document.

Member of the HIT2GAP consortium	Main role in the project	Role in Data Management Activities
<p>HIT2GAP consortium</p> <p>Each member of the consortium manipulating data in the frame of project activities</p>	<p>Data generation and management</p> <p>Communication and dissemination</p> <p>Exploitation</p> <p>Assurance quality control</p>	<ul style="list-style-type: none"> • Awareness and implementation of the DMP; • Is responsible for informing the dissemination manager of publications or patents being published in the frame of HIT2GAP; • Publishes data used during the development phase of HIT2GAP modules in an open access platform.

3.2 Data management process

This version of the D1.8 document is the second version of the DMP updated after 18 months of project and serves as a basis document for the implementation of the DMP.

This document will be updated as progresses are made in the project in terms of generated results, definition of exploitation associated to these results, and publications delivered during the project. To achieve this, the coordinator along with the innovation manager and the dissemination manager will ensure that data management is correctly addressed at all levels of the project (PMT level, WP level, task level, pilot site level...).

The Coordinator, along with the Innovation and Dissemination managers will identify annually the datasets that will be generated through deliverables or technical developments and which of them can be considered as Open Access. They will conduct discussions with the partners responsible for the deliverables and for the datasets that are generated.

An updated status of the data management within HIT2GAP project will be delivered as a reserved section in each periodic report and a dedicated section will be added in each project management plan.

3.3 Categories of datasets

In the frame of the HIT2GAP project, several categories of data and datasets can be defined:

- **Reserved or private data:** these data are used by the primary users (i.e., data owners),
- **Accessible and confidential data:** these data are used by the consortium members,
- **Available and public data:** these data are made available for third party/public users.

These data can be sparse in their nature. In this way, they can be grouped in two families: information data (for instance reports, deliverables, questionnaires related to the requirements of the different partners...) and technical data (e.g. measurements, IFC files, data produced by partners...).

3.4 Repository options and selection

For each category identified in Section 3.3, a storage repository should be selected to store the data in a way consistent with their open nature or not.

- **Reserved or private data:** the primary users will store these specific data on the servers of the company or institution to which they belong. This process is out of scope of the HIT2GAP DMP.
- **Accessible and confidential data:** some of these data will be stored in the EMDESK platform². The Coordinator has paid for an EMDESK license and has provided each partner with a secured and controlled access to the platform. EMDESK will be used all along the project and at the end of the project it will be decided with all the project members how the data stored in the EMDESK platform will be handled. The technical data (data collected in the pilot sites, data coming from the end-users...) will be stored in a dedicated storage repository. The retained storage solutions will depend on each pilot policy in terms of data open access. In case technical issues arise for the transfer of documents through EMDESK (e.g. voluminous data...), other solutions will be evaluated.
- **Available and public data:** all data will be made available on an Open Access repository, which will be Zenodo³ by default for the HIT2GAP beneficiaries. HIT2GAP is a “community” within ZENODO where all HIT2GAP papers should be linked. (<https://zenodo.org/communities/hit2gap/about/>, (see Figure 1). Other repositories can be used by partners as far as they offer similar accessibility assets and answer the European Commission requirements. Once data are stored, they will be bound to the HIT2GAP project through the OpenAire⁴ platform. This includes all data used in the project if not protected by a confidentiality or security clause, ruled by protection of personal data, or any other clause that would conflict with the publication of data.

² <https://EMDESK.eu>

³ <http://zenodo.org>

⁴ <https://www.openaire.eu/>

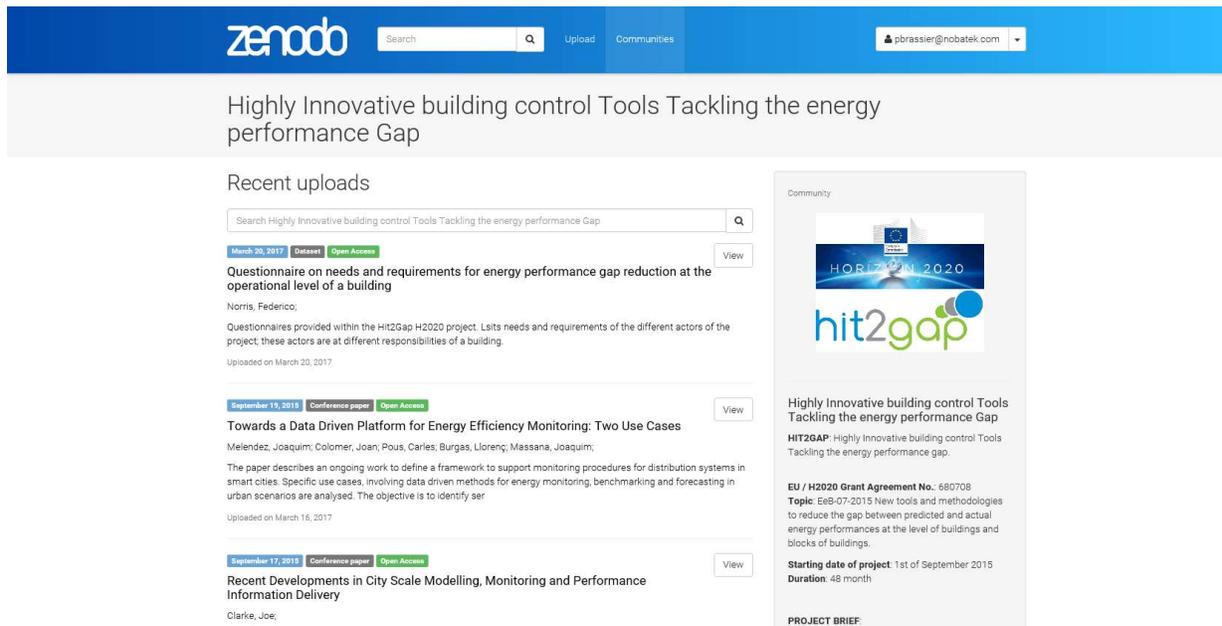


Figure 1: Zenodo "HIT2GAP" community

3.5 Lifecycle of the Data management

3.5.1 Dataset management template

The sequential management of the data generated or collected within HIT2GAP is organized as follows:

- The data are first identified and classified according to the three categories defined previously;

The data are then prepared in order to fill in the H2020 data management template presented below in

- Table 3;
- The data is stored in its corresponding repository (cf. previous section);
- The storage is then managed on the long-term (implementation of back-up storage solutions, accessibility management, data destruction...);

Table 3: H2020 Dataset template

H2020 Dataset Management Template	
Data set reference and name	Identifies the data set to be produced
Data set description	Description of the data that will be generated or collected, its origin, nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse
Standards and metadata⁵	Reference to existing suitable standards of the discipline that govern data collection, aggregation, storage and sharing.
Data sharing	<p>Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).</p> <p>In deciding on data sharing, ethical considerations, rules of personal data, intellectual property, commercial, privacy-related, and security-related aspects should be considered.</p>
Archiving and preservation (storage/backup):	Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.

⁵ **Metadata:** Metadata is “data about data” or “information about information.” It is an information layer assigned to publications or datasets that make them identifiable, linkable and searchable.

3.5.2 Generalities on data workflow in the HIT2GAP platform

The data collected and generated that are described in the present document include the data collection used in the global workflow of the platform developed in the HIT2GAP project. Various aspects are to be described to contextualize these data. The Table 4 provides a technical classification of these data (i.e. data to be considered in the global workflow of the platform).

Table 4: Technical classification of collected and generated data

Type of data		Description	
Collected data	Building-related	Dynamic	Various data will be measured on-sites, related to the building's working. Typically, such data are extracted from a BMS, or transmitted directly by sensors/actuators. Such data can be temperatures, humidity levels, state of a window opening, etc. Some measures will also be performed through similar systems regarding the external weather conditions.
		Static	Static data such as the description of a building will also be collected, through different methods, typically questionnaires, BIM files, drawings, operations & maintenance manuals, etc.
	Occupant-related	Dynamic	Data will be measured through wearable devices and positioning systems, to obtain information about an occupant's state (stress status for instance), and the global occupancy of specific zones in a building
		Static	A questionnaire will be transmitted to the occupants on a voluntary basis to collect their satisfaction level regarding energy-related comfort in a building
Generated data	Short term analysis	Predicted values for energy consumption, indoor temperatures, etc. Reports; Types of faults; Alerts; Source of error, etc. Recommendations; Actions; User State, etc.	
	Building performance simulation	Simulated values; reports; models	
	Visualization modules	NA	

One should note that all the data presented in Table 4 are part of the data that will be shared by the partners' modules in the frame of the HIT2GAP project. In this context, all these data will be stored

in the data storage solution implemented in the Core Platform, and available as inputs for every partners' modules.

The way these data are managed within the different software bricks is addressed in the documents related to ethics issues management (D9.1 [3], D9.2 [4], D9.3 [5] and D9.4 [6]).

4 Data set reference, name and description

This section describes the data that will be collected or generated during the HIT2GAP project and provides an identifier for the data set, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication.

This section provides also some information about the way these data will be managed (made accessible or not).

This section also provides the data management template that will be used within the HIT2GAP project and lastly it gives the template filled in with information related to the data collected in the frame of the project and identified so far.

4.1 Collected data

Table 5 gives a macro-categorization of the different data that will be collected

- On-site generated data collected and stored in the HIT2GAP platform;
- Data collected by the project partners in order to develop the project work.

Such categorization may be updated in future versions of the Data Management Plan based on the agreement of pilot sites to open access of the collected data.

Table 5: Data collected in the frame of the HIT2GAP project

Type/Category	Initial Identification	Accessibility /Sharing Strategy
Pilot data	Measurements conducted in the 4 pilot sites coming from BMS or other sources of measurement.	<ul style="list-style-type: none"> • Open by default • Restrictions can be made based on agreement of each pilot site.
	Questionnaire responses related to technical information about the buildings that include technical design drawings, technical data about the systems installed in the buildings, pilot site baseline situation.	<ul style="list-style-type: none"> • Open by default • Restrictions can be made based on agreement of each pilot site.
	Responses to the questionnaire submitted to the occupants of the pilot sites in order to collect their preferences.	<ul style="list-style-type: none"> • Confidential to consortium partners
Research and development data	Questionnaire responses related to needs and users' requirements	Open Access (the answers to the questionnaire have been uploaded on ZENODO, https://zenodo.org/record/401009#.WM_wb2_hCpo)
	Responses to the technical questionnaire submitted in order to collect the information related to the tools brought by each partner in the project	<ul style="list-style-type: none"> • Confidential to consortium partners

Market exploitation data	Exploitable results identified by each partner to feed the deliverable D6.1	<ul style="list-style-type: none"> Confidential to consortium partners.
--------------------------	---	--

4.2 Generated data

Table 6 gives a macro-categorization of the data that will be generated by the project.

Table 6: Technical data generated through the development and use of the tools developed within HIT2GAP

Type/Category	Initial Identification	Accessibility /Sharing Strategy
HIT2GAP internal data	Data models defined as part of the project for the development of the platform (WP1). Building Performance Dictionary	<ul style="list-style-type: none"> When consolidated, these data will be made available through the deliverables dedicated to the outcomes of the project; The data model and the dictionary can provide a set of metadata describing the HIT2GAP data. Possibility of associating an open source license to the data model and part of the core platform will be studied within the consortium as part of the business model elaboration.
	Software modules	<ul style="list-style-type: none"> When planned by partners, some software bricks or modules will be licensed under open source license, or with free access to use, according to the exploitation plan; Open access software can be available through a web repository: Github is a possibility for open source code.
HIT2GAP processed data	Pre-processed data of the pilots; Knowledge based (semantic level applied on the field level data)	<ul style="list-style-type: none"> When the platform will be available, these data can be made open for the public based on agreement of each pilot site; An access policy will be implemented and enforced for each specific use case in order to control and secure data access. Attention will be put on the privacy protection for occupants-related data.
	Post-processed and visualised data Results of the simulations applied to the pilot sites	<ul style="list-style-type: none"> When consolidated, these data can be made available through the deliverables dedicated to the outcomes of the project An access policy will be implemented and enforced for each specific use case in order to control and secure data access

4.3 Dataset management template

As previously stated, a DMP requires a data management template in order to clearly identify the considered dataset and provide a management procedure for each of them. The H2020 dataset management template provided in

Table 3 has been customised for the HIT2GAP project needs.

The following table gives the data management template that will be used in the frame of the HIT2GAP project.

Table 7: Data management template used in HIT2GAP

HIT2GAP Dataset Management Template	
WP / Task & Data Manager	
Dataset reference / name	
Mandatory Metadata	European Union H2020 HIT2GAP GA680708
Dataset Specific Metadata <i>(keyword(s) that categorize data to make it linked/searchable)</i>	
Data set description <i>(data description, origin, nature, scale, if it underpins a publication, who useful to, existence of similar data, possibilities for reuse)</i>	
Standards <i>(reference to existing standards in topic area governing data collection, aggregation, storage and sharing)</i>	
Data sharing <i>(how data will be shared, identification of repository, existence of embargo period if any, identification of software or tools necessary for reuse)</i>	
Archiving and preservation (storage/backup): <i>(procedure for long-term preservation, length of preservation, estimation of costs and how covered)</i>	

4.4 Datasets identified within HIT2GAP

This section details the data management templates filled in for data which are collected in the frame of the project and mentioned at the beginning of the present chapter (section 4.1). This part of the DMP will be the main part that will evolve during the lifespan of the project and will be updated and developed in a dynamic manner.

4.4.1 Building-related: measurements

HIT2GAP Dataset Management – Building-related measurements	
WP / Task / Data Manager	WP5 / MOS, BES, GIR, CYL, WRS
Dataset reference / name	Measurements conducted in the 4 pilot sites coming from BMS or other sources of measurement (individual sensors) except from individual devices.
Mandatory Metadata	European Union H2020 HIT2GAP GA680708
Dataset Specific Metadata <i>(keyword(s) that categorize data to make it linked/searchable)</i>	Energy consumptions Water consumption Indoor comfort conditions Indoor temperature Indoor relative humidity External weather data Physical data of energy systems (air temperature after heat recovery, Supply air temperature, Hot water supply temperature, pressure, Heat flow...)
Data set description <i>(data description, origin, nature, scale, if it underpins a publication, who useful to, existence of similar data, possibilities for reuse)</i>	These data are collected from the pilot sites through BMS installed in the building or from additional measurement devices that may be installed to make the data more complete and more accessible to the project. Presented as time series data. These data will be used for calibration of the simulation models, and as an input of modules in the management level (forecasting, FDD, behaviour modelling, energy management system, and visualization modules)
Standards <i>(reference to existing standards in topic area governing data)</i>	The data can be presented in CSV, XML or JSON format, or any other non-proprietary simple format for time-series data.

HIT2GAP Dataset Management – Building-related measurements	
<p><i>collection, aggregation, storage and sharing)</i></p>	<p>They can also be stored in any data storage solution. In the HIT2GAP project, the data storage solution brought by FISE is likely to be used; it is technically built upon HDF5 format, a data-storage solution oriented to large time-series data.</p>
<p>Data sharing <i>(how data will be shared, identification of repository, existence of embargo period if any, identification of software or tools necessary for reuse)</i></p>	<p>Once the platform is up and running, the data that can be collected from the pilots and alternative sources will be made available for public access based on agreements of each pilot site.</p> <p>If not too voluminous, these data will be stored by default on Zenodo. If too voluminous, the consortium will identify another suitable open access dataset repository.</p>
<p>Archiving and preservation (storage/backup): <i>(procedure for long-term preservation, length of preservation, estimation of costs and how covered)</i></p>	<p>For the long-term storage, the data can be stored on a hard drive if needed. This should be discussed when the data collection will start.</p> <p>Published data will be stored for at least 4 years after the end of the project.</p>

4.4.2 Building-related: description

HIT2GAP Dataset Management - Building-related description	
WP / Task / Data Manager	WP5 / MOS, BES, GIR, CYL, WRS, NUIG
Dataset reference / name	<ol style="list-style-type: none"> 1. Questionnaire responses related to technical information about the buildings, technical design drawings, pilot site baseline situation 2. BIM files
Mandatory Metadata	European Union H2020 HIT2GAP GA680708
Dataset Specific Metadata <i>(keyword(s) that categorize data to make it linked/searchable)</i>	Drawings Energy bills Descriptions of the systems installed in the buildings Building envelop information HVAC systems information Descriptions of Energy sources Heating/cooling set point Schedules Zones for heating Cooling regulations Lighting systems information Building control infrastructure Building occupants' information Location of the building
Data set description <i>(data description, origin, nature, scale, if it underpins a publication, who useful to, existence of similar data, possibilities for reuse)</i>	<p>The data contains the responses to the questionnaires sent by MOS to the pilot site managers at the beginning of the project. These data are used for the simulations of the energy performances of the pilots, as inputs of the fault detection and diagnosis (FDD) modules and the 3D visualization module.</p> <p>These data, when possible, will be directly collected from a BIM file.</p>
Standards <i>(reference to existing standards in topic area governing data collection, aggregation, storage and sharing)</i>	<ol style="list-style-type: none"> 1. Questionnaire: there are no specific standards applicable for these data sets. 2. IFC is an open format for BIM files; IFC is a standard format that provides with an open and interoperable data schema for exchangeable

HIT2GAP Dataset Management - Building-related description	
	BIM models. IFC is curated by the building smart organization (buildingsmart.org).
Data sharing <i>(how data will be shared, identification of repository, existence of embargo period if any, identification of software or tools necessary for reuse)</i>	<p>The questionnaire answers will be stored on EMDESK. Files will be provided in pdf format or any other non-proprietary format.</p> <p>IFC files will be stored in EMDESK if supported, as IFC files can be voluminous; otherwise there will be stored on local hard drives or in a cloud solution internal to the project.</p> <p>If authorization is provided by pilot site managers, data will be made available on an open access repository. The possibility of having an open access for only part of the data is a possibility depending on the criticality of each data.</p>
Archiving and preservation (storage/backup): <i>(procedure for long-term preservation, length of preservation, estimation of costs and how covered)</i>	<p>Project data produced by the partners involved will be saved and backed up on the specific partners' own backup system, and on EMDESK.</p> <p>Published data will be stored for at least 4 years after the end of the project.</p>

4.4.3 Users-related: measurements

HIT2GAP Dataset Management – User-related measurements	
WP / Task / Data Manager	WP5 / MOS, BES, GIR, CYL, WRS WP3 / EGE, EUR, TEK
Dataset reference / name	Measurements conducted in the 4 pilot sites coming from smart devices and positioning solutions.
Mandatory Metadata	European Union H2020 HIT2GAP GA680708
Dataset Specific Metadata <i>(keyword(s) that categorize data to make it linked/searchable)</i>	Buildings/Zones/Rooms occupancy Occupants status (cardiac rhythm)
Data set description <i>(data description, origin, nature, scale, if it underpins a publication, who useful to, existence of similar data, possibilities for reuse)</i>	These data are collected from the pilot sites through specific systems: a positioning system, and smart devices (smart watch, smart phone...) worn by occupants on a voluntary basis. Presented as time series data. These data will be used for the modelling of behaviour.
Standards <i>(reference to existing standards in topic area governing data collection, aggregation, storage and sharing)</i>	The data can be presented in CSV, XML or JSON format, or any other non-proprietary simple format for time-series data. They can also be stored in any data storage solution. In the HIT2GAP project, the data storage solution brought by FISE is likely to be used; it is technically built upon HDF5, a data-storage solution oriented to time-series data.
Data sharing <i>(how data will be shared, identification of repository, existence of embargo period if any, identification of software or tools necessary for reuse)</i>	These data will be confidential to consortium partners and stored on EMDESK.
Archiving and preservation (storage/backup): <i>(procedure for long-term preservation, length of preservation, estimation of costs and how covered)</i>	For the long-term storage, the data can be stored on a hard drive if needed. This should be discussed when the data collection will start. In the frame of the project, these data will be stored in a HDF5 solution which offers functionalities for backup. Published data will be stored for at least 4 years after the end of the project.

4.4.4 Users-related: preferences

HIT2GAP Dataset Management – User-related preferences	
WP / Task / Data Manager	WP5 / MOS, BES, GIR, CYL, WRS WP3 / EGE
Dataset reference / name	Responses to the questionnaire submitted to the occupants of the pilot sites in order to collect their preferences
Mandatory Metadata	European Union H2020 HIT2GAP GA680708
Dataset Specific Metadata <i>(keyword(s) that categorize data to make it linked/searchable)</i>	Workspace Comfort (air quality, temperature) measured through satisfaction level, perception Occupancy Features adjustment Preference Ambient factors
Data set description <i>(data description, origin, nature, scale, if it underpins a publication, who useful to, existence of similar data, possibilities for reuse)</i>	The data contains the responses to the questionnaires prepared by EGE in order to develop occupant behaviour model. The answers are specific to the behaviour modelling module and hardly reusable.
Standards <i>(reference to existing standards in topic area governing data collection, aggregation, storage and sharing)</i>	There are no specific standards applicable for these data sets.
Data sharing <i>(how data will be shared, identification of repository, existence of embargo period if any, identification of software or tools necessary for reuse)</i>	The questionnaire answers will be confidential to the consortium and stored on EMDESK. The files will be provided in pdf format or any other non-proprietary format.
Archiving and preservation (storage/backup): <i>(procedure for long-term preservation, length of preservation, estimation of costs and how covered)</i>	Project data produced by the pilots involved will be saved either by the partner responsible for this task or the pilot sites or the HIT2GAP platform. Published data will be stored for at least 4 years after the end of the project.

4.4.5 Generated by modules

HIT2GAP Dataset Management – Data generated by modules	
WP / Task / Data Manager	WP5 / MOS, BES, GIR, CYL, WRS
Dataset reference / name	Data generated by short-term analysis modules (FDD, forecasting, behaviour modelling) and simulation modules
Mandatory Metadata	European Union H2020 HIT2GAP GA680708
Dataset Specific Metadata <i>(keyword(s) that categorize data to make it linked/searchable)</i>	Predicted consumptions, temperatures Alerts Recommendations Simulated values Reports
Data set description <i>(data description, origin, nature, scale, if it underpins a publication, who useful to, existence of similar data, possibilities for reuse)</i>	These data are generated by partners' modules, based on inputs which are either user-related data, building-related data, or other generated data. Some data can be presented as time series data (forecasted data, simulated data). The amount data can be massive depending on the frequency at which data are produced. These data are to be used by other modules (simulation modules, short-term analysis modules, or visualization modules)
Standards <i>(reference to existing standards in topic area governing data collection, aggregation, storage and sharing)</i>	Time-Series data can be presented in CSV, XML or JSON format, or any other non-proprietary simple format. There is no standard for the other data. Nevertheless, these data will be collected in the core platform through web API that will use JSON or XML as a communication language. These data will be stored in the data storage of the Core Platform: as RDF triples or in the HDF5 solution depending on their nature.
Data sharing <i>(how data will be shared, identification of repository, existence of embargo period if any, identification of software or tools necessary for reuse)</i>	Once the platform is up and running, the data generated by the modules will be made available for public access based on agreements of each pilot site and module developer. If not too voluminous, these data will be stored by default on Zenodo. If too voluminous, the consortium

HIT2GAP Dataset Management – Data generated by modules	
	will identify another suitable open access dataset repository.
Archiving and preservation (storage/backup): <i>(procedure for long-term preservation, length of preservation, estimation of costs and how covered)</i>	<p>For the long-term storage, the data can be stored on a hard drive if needed. This should be discussed when the data collection will start.</p> <p>Published data will be stored for at least 4 years after the end of the project.</p>

4.4.6 Internal: needs and users requirements

HIT2GAP Dataset Management – Needs and requirements of end-users	
WP / Task / Data Manager	WP1 / NBK
Dataset reference / name	Questionnaire responses related to needs and users requirements
Mandatory Metadata	European Union H2020 HIT2GAP GA680708
Dataset Specific Metadata <i>(keyword(s) that categorize data to make it linked/searchable)</i>	BMS providers' needs Facility managers' needs System manufacturers needs Exploitation and maintenance companies' needs Building owners needs Building managers needs Users' needs Energy performance and efficiency
Data set description <i>(data description, origin, nature, scale, if it underpins a publication, who useful to, existence of similar data, possibilities for reuse)</i>	The dataset is composed of the records of the stakeholders' answers to the questionnaire/interview and the literature review. These data will be used to underpin the modules definition within WP1 as well as the services proposition in response to the market needs (WP6).
Standards <i>(reference to existing standards in topic area governing data collection, aggregation, storage and sharing)</i>	There are no specific standards applicable for these data sets
Data sharing <i>(how data will be shared, identification of repository, existence of embargo period if any, identification of software or tools necessary for reuse)</i>	These data are fully open. The answers to the questionnaire have been uploaded on ZENODO, https://zenodo.org/record/401009#.WM_wb2_hCpo .
Archiving and preservation (storage/backup): <i>(procedure for long-term preservation, length of preservation, estimation of costs and how covered)</i>	Project data produced by the partners involved will be saved and backed up on the specific partners' own backup system.

4.4.7 Internal: description of partners' software in HIT2GAP project

HIT2GAP Dataset Management – description of partners' software	
WP / Task / Data Manager	WP1 / API
Dataset reference / name	Questionnaire responses related to the tools brought by each partner in the project
Mandatory Metadata	European Union H2020 HIT2GAP GA680708
Dataset Specific Metadata <i>(keyword(s) that categorize data to make it linked/searchable)</i>	Modules/Algorithms/Data treatment tool/Forecasting tool/Fault Detection & Diagnosis tool/ Data collection tool / Visualisation tool/user behaviour tool/Energy management tool/3D model/End-user type/ use case/input data/output data
Data set description <i>(data description, origin, nature, scale, if it underpins a publication, who useful to, existence of similar data, possibilities for reuse)</i>	The dataset is composed of the responses to template provided to the partners bringing an existing tool to the HIT2GAP project.
Standards <i>(reference to existing standards in topic area governing data collection, aggregation, storage and sharing)</i>	There are no specific standards applicable for these data sets.
Data sharing <i>(how data will be shared, identification of repository, existence of embargo period if any, identification of software or tools necessary for reuse)</i>	The questionnaire answers is stored on EMDESK. The files are in pdf format.
Archiving and preservation (storage/backup): <i>(procedure for long-term preservation, length of preservation, estimation of costs and how covered)</i>	Project data produced by the partners involved can be saved and backed up on the specific partners' own backup system or the HIT2GAP storage based on each case.

4.4.8 Internal: Market exploitation data

HIT2GAP Dataset Management – description of partners' software	
WP / Task / Data Manager	WP6 / R2M
Dataset reference / name	Exploitable results identified by the partners
Mandatory Metadata	European Union H2020 HIT2GAP GA680708
Dataset Specific Metadata <i>(keyword(s) that categorize data to make it linked/searchable)</i>	Exploitable results
Data set description <i>(data description, origin, nature, scale, if it underpins a publication, who useful to, existence of similar data, possibilities for reuse)</i>	The dataset is composed of the list of exploitable results defined by each partner and the associated SWOT analysis.
Standards <i>(reference to existing standards in topic area governing data collection, aggregation, storage and sharing)</i>	There are no specific standards applicable for these data sets.
Data sharing <i>(how data will be shared, identification of repository, existence of embargo period if any, identification of software or tools necessary for reuse)</i>	The D6.1 is aggregating the exploitable results defined by the partners. This is a confidential document. Therefore these data will remain confidential to the consortium.
Archiving and preservation (storage/backup): <i>(procedure for long-term preservation, length of preservation, estimation of costs and how covered)</i>	Project data produced by the partners involved will be saved and backed up on the specific partners' own backup system.

5 Conclusions

This deliverable has provided details about the data management plan envisioned within the HIT2GAP project. This is the second version of the DMP delivered after 18 months of project.

The main contributors in the data management of the project have been identified and their roles have been defined in accordance with the management structure of the project.

Open access consideration has also been introduced in the document in relation to the data collected and generated as part of the HIT2GAP project. In this respect:

- Scientific publications and related datasets will be published with an open access. The proposed methodology is:
 - To grant open access to all project results in the ways described in this document;
 - To upload publications and datasets in an open repository, Zenodo by default;
 - To reference publications and datasets with HIT2GAP in the OpenAire platform;
 - Data will be available under an open access whenever this choice is not in conflict with confidential, security or privacy clauses. Discussions will be made with the different partners involved.
- Confidential datasets are internal to the consortium and will be uploaded in the EMDESK platform by default;
- Private data will be stored in local machines and will not be shared within the consortium.

A first list of datasets collected and generated within the HIT2GAP project has been elaborated, and the way these data will be available has been discussed. A special attention has been put on using standard open access modes to the publication and data that will be released.

Internal guidelines have been elaborated to give all project beneficiaries the description of the process retained for the data open access within the HIT2GAP project.

This DMP will be updated in parallel with periodic reports and project management plans. The innovation manager, communication manager and project coordinator serve as references for questions related to data management in HIT2GAP.

6 References

- [1] H2020 Programme - Guidelines on FAIR Data Management in Horizon 2020, Version 3.0, 26 July 2016. Available at http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.
- [2] H2020 Programme - Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, Version 3.2, 21 March 2017. Available at http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.
- [3] D9.1 “NEC-Requirement n°3”, HIT2GAP project, March 2017.
- [4] D9.2 “POPD – Requirement n°2”, HIT2GAP project, March 2017.
- [5] D9.3 “H – Requirement n°1”, HIT2GAP project, March 2017.
- [6] D9.4 “M – Requirement n°4”, HIT2GAP project, March 2017.

7 Glossary and acronyms

API	APINTECH
BES	Bouygues Energies & Services
BMS	Building Management System
BRE	Building Research Establishment
CC	Creative Commons
CYL	Cylon Controls
DM	Data management
DMP	Data Management Plan
EGE	Ege University
EUR	EURECAT
FISE	Fraunhofer ISE
GIR	Giroa
GM	General Meeting
H2020	Horizon 2020 research and innovation programme
IPR	Intellectual Property Rights
LIU	LIUPPA Université de Pau et des Pays de l'Adour
MOS	MOSTOSTAL
NBK	NOBATEK
NUIG	University of Galway
OA	Open Access
PMT	Project Management Team
TEK	IK4-Tekniker
WP	Work Package
WRS	City of Warsaw

8 Annex: Guidelines related with the process for open access data within HIT2GAP

8.1 H2020 Open Access policy

This document is aimed at clarifying H2020 policy on open access and diffusion of project results and dissemination activities. All project beneficiaries are obliged under the H2020 Grant Agreement (Art. 29.2) to ensure **open access** to all peer-reviewed scientific publications and (Art. 29.3) research data;

How to comply?

Compliance with this open access policy is achieved by:

1. Granting open access to publications and research data and;
2. Uploading publications and/or data to public repositories.

In addition, all partners need to consider that:

1. The costs associated with publications should be handled by the partner who generates the publications. These costs are eligible so they can be declared to the EC (for instance in relation with the dissemination budget). Partners are responsible for managing their own budget. Please keep in mind that the budget is not extensible.
2. All beneficiaries must give 45 days advance notice to other project beneficiaries about the intention of disseminating results (Art. 29.1);
3. All beneficiaries should contact HIT2GAP Task-7.3 (luismiguel.blanesrestoy@nuigalway.ie) prior to submitting your results in order follow the correct procedures and to centralize curation of all HIT2GAP repositories and HIT2GAP reporting activities;
4. All beneficiaries must include the following disclaimer in dissemination and publications: "This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 680708" in order to properly link them to H2020 HIT2GAP project in repositories.

8.2 Open Access Options

Regarding peer-reviewed journal publications the following options are available:

1. Publications are published in **open access** journals. This choice is encouraged to all project partners. This means the journal grants open access by definition, these are normally:
 - a. Conference proceedings (IBPSA, CIBSE Technical Symposium, etc.);

- b. Open access journals like ITCon (<http://www.itcon.org/about>) with no Article Processing Charges (APC);
 - c. Open access journals like Cogent Engineering with "low cost" APCs (<https://www.cogentoa.com/journal/engineering>) by Taylor&Francis
2. Publications are published in **gold open access mode**, that means a higher fee is paid for APCs and the publication is granted open access immediately;
 3. Publication are published in **green open access mode**. That means the publication can be made available after an embargo period, which is set up by the publisher. The publication fee is lower that gold open access, or it might be funded by library subscription fees. For all the journals consulted, the embargo period is substantially for a longer period (18 months, 24 months etc.) than what is required under H2020 (6 months). Therefore, this option is not workable for H2020 projects;
 4. Publications are **not published in open access mode**. This options is not encouraged.

8.3 Open access repositories

For HIT2GAP participants the default repository to upload scientific dissemination material, is [ZENODO](https://zenodo.org). HIT2GAP is a "community" within ZENODO where all HIT2GAP papers should be linked. ([https://zenodo.org/communities/hit2gap/about/.](https://zenodo.org/communities/hit2gap/about/))

Zenodo is repository for both scientific dissemination articles and research data. Please remember that all peer-reviewed participation in conferences need to be considered in this repository.

The repository works very straightforward; you can use your ORCID account ID or use an email to register as a user and upload a paper. An automatic Creative Commons license generator is provided.

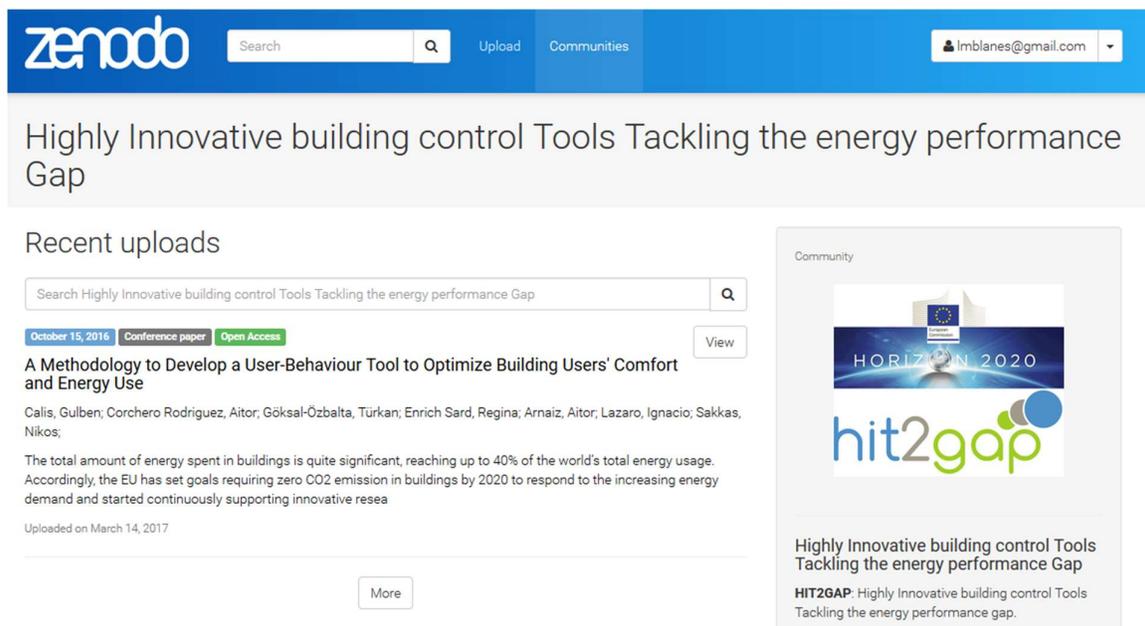


Figure 2 - Zenodo "HIT2GAP" community

When uploading a publication, participants should fill in the “funder” and “communities” fields with “hit2gap” and “OpenAIRE” text so zenodo can show look for the correct tag and index the publication to HIT2GAP. For best results please contact: luismiquel.blanesrestoy@nuigalway.ie.

8.4 Open source licensing options

The Creative Commons (CC) license is a convention to let share, distribute, modify and acknowledge your work free and openly as an alternative to complicated application of intellectual property rights. In order to apply CC, you just need to:

1. **Choose the CC license** most appropriated type using the license chooser [here](#). There are 7 types of licenses (combinations of 4 properties). The most appropriate one for most articles is the "Attributions-NoDerivs CC BY-ND". The different types of licenses can be consulted [here](#), with examples [here](#).
2. **Marking your work.** This can be done in a number of ways, from copying and pasting a disclaimer and logo, to embedding a piece of code in a web page. More details for good practice [here](#).

The [license](#) itself is a legal type of text organized by articles. There is also a friendlier short description ([Commons Deeds](#)), and an electronic machine readable version. The B2SHARE portal offers a [license selector](#) , but this is only a portal to share data.



Figure 3 – Creative Commons “Attributions-NoDerivs CC BY-ND” license

8.5 Resources

On EMDESK:

Document Manager: “EMDESK/Work/WP7.../H2020 Open Access Supporting Material”

Online:

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

https://ec.europa.eu/programmes/horizon2020/sites/horizon2020/files/FactSheet_Open_Access.pdf
