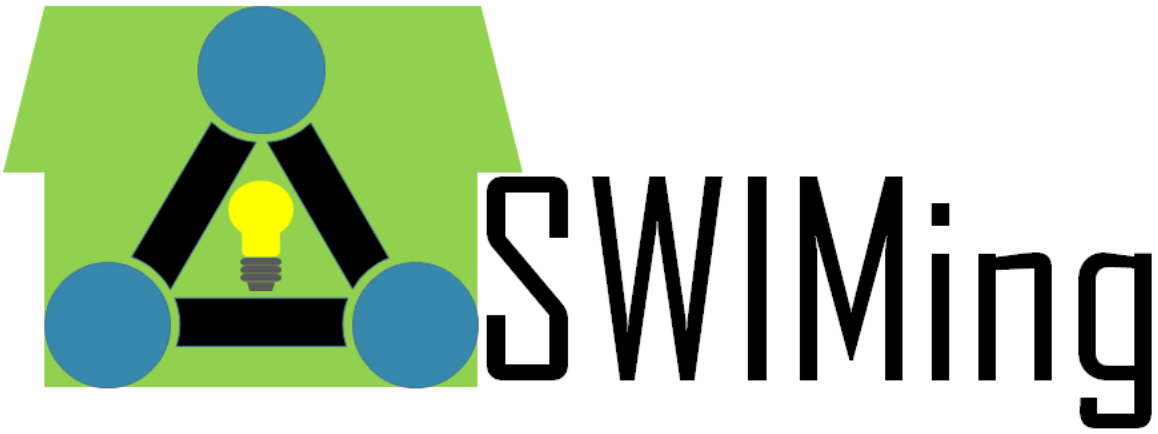

SWIMING: H2020 - 637162

Semantic Web for Information Modelling in Energy Efficient Buildings

Deliverable number **D2.1**

Deliverable title **Data Management Plan**

Main Authors **Matthias Weise, Kris McGlinn**

Grant Agreement number	637162
Project ref. no	H2020 - 637162
Project acronym	SWIMing
Project full name	Semantic Web for Information Modelling in Energy Efficient Buildings
Starting date (dur.)	01/02/2015 (24 months)
Ending date	31/01/2017
Project website	www.swiming-project.eu

Coordinator	Kristian McGlinn
Address	F 35, Computer Science, Trinity College Dublin
Reply to	Kris.McGlinn@scss.tcd.ie
Phone	+353 (0)1 896 8431
Document Identifier	D2.1

Class Deliverable	SWIMing EU-EEB-2015-637162
Version	v1
Document due date	31/07/2015
Submitted	29/07/2015
Responsible	Matthias Weise (AEC3)
Reply to	mw@aec3.de
Document status	Final
Nature	R(Report)
Dissemination level	PU(Public)
WP/Task responsible(s)	WP2/T2.2
Contributors	Matthias Weise (AEC3), Kris McGlinn (TCD), Hendro Wicaksono (KIT), Nikolaos Kaklanis (CERTH), Dimitrios Tzovaras (CERTH), Ioanna Petri (CERTH), Willie Lawton (Tyndall)
Distribution List	Consortium Partners
Reviewers	Hendro Wicaksono (KIT)
Document Location	swiming-project.eu/?page_id=7

Executive Summary

The SWIMing project is participating in the H2020 Pilot on Open Research Data, which aims to maximise access to and re-use of the research data generated during the course of the project. Towards this goal the following Data Management Plan (DMP) is presented. The DMP describes the full data management life cycle for all data sets that are collected, processed or generated over and beyond the duration of the SWIMing project. The DMP sets out to document the types of data collected, processed or generated during the project, methodologies and standards used, how that data will be exploited and made accessible for verification and re-use, and finally, how the data will be curated and preserved.

The presented DMP is implementing the “Guidelines on Data Management in Horizon 2020” (2013) and follows the new W3C draft “Data on the Web Best Practices” released in June 2015. Both guidelines are used as a reference to explain the implementation of the DMP in the SWIMing CSA project. The W3C guideline identifies twelve challenges for publishing data in the web, which are discussed in the context of SWIMing, in particular the use cases that are collected, harmonized and enriched throughout the project work. Not all challenges are equally important due to the descriptive nature of our data, which in fact is mainly on the level of metadata (data about data) and not on the level of energy related building data like for instance all kinds of measurements delivered by sensor networks. Accordingly, additional information is provided in the deliverable not only about our data but also about the predominantly manual data collection method. The data collection itself is an iterative process that already went through several steps of refinement and still is under development. This is an ongoing effort that was kicked-off by the SWIMing project and is to be continued by the Linked Building Data community.

Data collection within SWIMing is already done in a broader context trying to establish relationships to other research (we have currently reviewed 38 existing EeB projects) and CSAs (EEBERs, AMANAC and EEB-CA2) projects as well as relevant standardization bodies, in particular W3C and buildingSMART being the key players for semantic web technologies and the specification of shared building data. The overall process is already controlled by a W3C community group and follows the IDM/MVD methodology for use case documentation developed by buildingSMART. All data is published on a Wiki page, which reflects the latest state of use case developments. Extensions towards a machine readable data representation are proposed but still need to be discussed and agreed. Also, the relationship to other data sources like the ontology and dataset catalogue developed by the Ready4SmartCities project need to be clarified in order to efficiently combine and integrate other results.

Document Information

IST Project Number	H2020 - 637162	**Acronym**	SWIMing
Full Title	Data Management Plan		
Project URL	www.swiming-project.eu		
Document URL	swiming-project.eu/?page_id=7		
EU Project Officer	Jose Riesgo		

Deliverable	**Number**	D2.1	**Title**	Data Management Plan
Workpackage	**Number**	WP2	**Title**	Data Management Plan

Date of Delivery	**Contractual**	31st July 2015	**Actual**	29th July 2015
Status	version 1		final □	
Nature	prototype □ report □ dissemination ■			
Dissemination level	public ■ consortium □			

Authors (Partner)	AEC3 , CERTH, TCD , KIT, UCC			
Responsible Author	**Name**	Matthias Weise	**E-mail**	mw@aec3.de
	Partner	AEC3	**Phone**	

Abstract (for dissemination)	This deliverable presents the data management plan in the SWIMing project (WP2).
Keywords	SWIMing, data management

Version	Modification(s)	Date	Author(s)
1	First draft	02/07/2015	Matthias Weise, Kris McGlinn, Hendro Wicaksono, Nikolaos Kaklanis, Dimitrios Tzovaras, Ioanna Petri, Willie Lawton
2	Version to be sent for peer-review	27/07/2015	Matthias Weise, Kris McGlinn, Hendro Wicaksono, Nikolaos

			Kaklanis, Dimitrios Tzovaras, Ioanna Petri, Willie Lawton
3	Submitted version	29/07/2015	Matthias Weise

Project Consortium Information

Participants		Contact
The Provost, Fellows, Foundation Scholars & The Other Members of Board of The College of the Holy & Undivided Trinity of Queen Elizabeth near Dublin (Trinity College Dublin, Ireland)		**Kris McGlinn** Kris.McGlinn@scss.tcd.ie
Institute for Information Management in Engineering		**Hendro Wicaksono** hendro.wicaksono@kit.edu
Centre for Research and Technology Hellas Information Technologies Institute		**Dimitrios Tzovaras** Dimitrios.Tzovaras@iti.gr
Tyndall National Institute, University College Cork		**Willie Lawton** willie.lawton@tyndall.ie
AEC3 Ltd.		**Matthias Weise** mw@aec3.de

Table of Contents

List of Abbreviations

Abbreviation	Definition
AEC	Architecture, Engineering and Construction
BIM	Building Information Modelling
BIM-LD	Building Information Modelling – Linked Data
BLC	Building Life Cycle
BLCEM	Building Life Cycle Energy Management
BPMN	Business Process Modelling Notation
CSA	Coordination & Support Action
CSV	Comma-Separated Values
DMP	Data Management Plan
EEB	Energy-efficient Buildings
EMS	Energy Management System
ER	Exchange Requirement
EU	European Union
FM	Facilities Management
GIS	Geographic Information System
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
IDM	Information Delivery Manual
IEC	International Electrotechnical Commission
IFC	Industry Foundation Classes
ISO	International Organization for Standardization
JSON-LD	JavaScript Object Notation for Linked Data
LBD	Linked Building Data
LD	Linked Data
LOD	Linked Open Data (in Semantic Web), Level of Detail/Development (in AEC industry)
MVD	Model View Definition
OWL	Ontology Web Language
RDF	Resource Description Framework
SQL	Structured Query Language
TCD	Trinity College Dublin
URI	Uniform Resource Identifier
URL	Uniform Resource Locator

1 Objectives

The objective of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the applicants with regard to the datasets that will be generated by the project. The DMP is a new important element in Horizon 2020 projects and describes what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved.

SWIMing is a Coordination & Support Action (CSA) and does not actively research on topics related to Energy-efficient Buildings (EeB) and the use of Linked Building Data (LBD). The aim of SWIMing is not to generate new data, but to review the types of domains, use cases and data modelling that EeB projects are addressing and identifying how Building Information Modelling - Linked Data (BIM-LD) technologies can support the exploitation of project results. Nonetheless, SWIMing will generate data in the form of business use cases, guidelines and best practices. This data should be publicly available, comparable, correct, up-to date, complete and compelling and ideally maintained by an active and neutral EeB community. A specific challenge of SWIMing is to extract and harmonize relevant data from very different project resources like project websites, deliverables, publications, tools and feedback from project partners. Such neutral knowledge base will foster reuse of project results and better collaboration, and help in the process of identifying common data requirements, which can benefit from the application of Linked Open Data (LOD) technologies.

This deliverable shows the approach that has been chosen by the SWIMing project to deal with expected project results. It first clarifies the types of managed data and the used methodology to collect and harmonize that data and then explains the way in which SWIMing is dealing with the challenges of data management and publication as mentioned in the Horizon 2020 Data Management guidelines [3] and also the W3C guidelines and best practices for managing data on the web [4]. It should be noted that the types of data SWIMing will generate do not necessarily subscribe to all the recommendations put down by the EC and W3C, but we address each guideline in respect to the data regardless.

2 Types of Managed Data in SWIMing

The SWIMing project, as a CSA, will collect data in the form of relevant business use cases in and around the different Building Life Cycle Energy Management (BLCEM) stages and Building Information Modelling (BIM) requirements for these use cases. It may also generate new business use cases which can benefit from the application of BIM-LD during the course of analyzing projects and liaising with academic, industrial and

governmental bodies. The project will also provide a set of guidelines and best practices for generating free interlinked, and semantically interoperable BIM resources for meeting current and future application requirements within the BLC, uncovered during the analysis of the business use cases. It will therefore generate a set of guidelines and best practices for:

1 Standardization of project outcomes through shared linked data vocabularies. Examples of these are: building system control data model, data models for communication between the building and the wider 'smart grid', models for describing new energy saving materials and devices, models of devices and sensors in terms of costs, energy ratings and their capabilities, models for describing occupant behavior and comfort, etc.

2 Minimizing time, cost and resources employed in integrating (reformatting, interlinking) existing EeB project outcomes into the BIM-LD cloud;

3 Generating and exploiting these BIM-LD outcomes to meet new and future application requirements;

4 Identifying and developing LD-based applications for frequent and common BIM related tasks.

The set of guidelines and best practices will be created/updated in each iteration, which will be put at the disposal of the Steering board, which has been created in WP4 and which consists of the project partners and also the W3C LBD community members, to allow them to contribute with further resources and use cases. Our purpose is to guide the transformation of such resources in a way that allows for their reuse and interoperation across the BLC and on the Web, by following Open Data standards.

The W3C community portal and wiki will be the main port of call for any community member to contribute to the development of the business use cases. Here they will also be able to contribute to the classification and categorization of stakeholders and data domains. They will be encouraged to share the data models and open data sets they use with the wider community.

The types of data generated on the wiki will therefore be use case descriptions, guidelines, and best practices. A full description of the organization of the use cases, domains, stakeholders can be found in D1.1 as well as on the shared wiki [2]. Also, on this wiki under data domains a collection is being iteratively generated of typical data models (both non-RDF and RDF based) currently being used by the projects. This data is community driven and already put under control of the W3C LBD community group, and as such not all use cases are necessarily of direct relevance to the EeB domain. This is because the W3C group is interested in all data generated across the BLC. Nonetheless, most of the use cases are energy related as SWIMing is currently the main

driver of use case contributions. More details on the guidelines will be available when D2.2 is made available in M11 of the project.

3 Data Collection Framework

As shown in the previous section a main outcome of SWIMing in terms of managed and published data is to identify EeB business use cases which can benefit from the application of both Building Information Modelling and Linked Open Data (BIM-LD). Various EeB research projects will be reviewed, categorized and brought together in order to facilitate knowledge sharing and to increase the impact of project results. A main challenge of this data collection process is to find a common methodology to describe and compare identified business use cases. Thus, to be able to identify similarities and differences a common framework is needed that enables to categorize and cluster business use case developments.

The non-profit organization buildingSMART is developing open standards for the AEC/FM industry supporting data sharing throughout the life-cycle of a building. The open IFC standard (ISO 16739) is a main driver for the implementation of the BIM approach and is an internationally accepted reference for vendor-neutral data exchange of building data. buildingSMART is faced with very similar challenges as the SWIMing project because tool vendors are not able to support the whole IFC standard. Instead, they implement subsets of IFC being relevant for their specific application area. For instance, the CAD application of an architect is typically not able to handle structural analysis data of the structural engineer, or the tool might be limited to the early design stage and does not support later detailed design. To be able to manage design processes based on use case specific tools and partial data exchange the IDM/MVD methodology has been developed by buildingSMART. This methodology has been adopted by the SWIMing project for the data collection process.

The IDM/MVD methodology defines how to specify business use cases and how to coordinate involved stakeholders with their tools and data requirements. A prerequisite for this is to be clear about processes, actors, shared or exchanged data and used interfaces or data structures. It provides a framework for the specification of collaborative design scenarios, in particular for Building Information Modelling (BIM). The next subchapters briefly introduce into the IDM/MVD methodology and the types of data that are collected from EeB projects.

3.1 IDM/MVD methodology and its adoption in SWIMing

The IDM/MVD methodology is divided into two main parts:

1 Information Delivery Manual (IDM, orange parts in Figure 1)

2 Model View Definition (MVD, blue parts in Figure 1)

3.1.1 Information Delivery Manual

The Information Delivery Manual method (IDM, [9]) is focusing on knowledge defined by domain experts. It defines processes and exchange requirements, which will answer what kind of tasks must be carried out, who is responsible, when they have to carry out (order, dependencies) and what data needs to be exchanged.

Two kinds of specifications are used:

1 Process Maps based on the Business Process Modelling Notation (BPMN)

2 Exchange Requirements typically collected in a table format

Process Maps define the various tasks to be carried out throughout the life-cycle of a building. Each task is placed within a swim lane, which is assigned to an actor role whole is responsible for carrying out those tasks. Arrows between tasks define data dependencies and are typically linked with data exchange requirements. For making data exchanges more explicit IDM introduces own swim lanes, which may carry additional information about the kind of data source like BIM, drawings, regulations or other kinds of data. The horizontal axis is tailored according to the life-cycle phases so that it is visible whether a task has to be carried out in the feasibility stage, early design, detailed design, commissioning, construction phase or other phases.

More details might be added to refine processes and deal with alternatives. For instance tasks might be subdivided into subtasks, decision gateways might be introduced to control the data flow and to deal with iterative design cycles, or messages are added to show expected communication between actors. For SWIMing this level of detail is not relevant as the main focus is to agree on actor roles (domains & stakeholder), the design phases (building life-cycle stages) and tasks (use cases).

Exchange Requirements specify the data that needs to be exchanged. As mentioned above it typically starts with identifying main data sources in terms of high-level data structures or domains. This information can be represented in own swim lanes and will be detailed in the next step in order to identify required data, which is defined by objects, attributes and relationships.

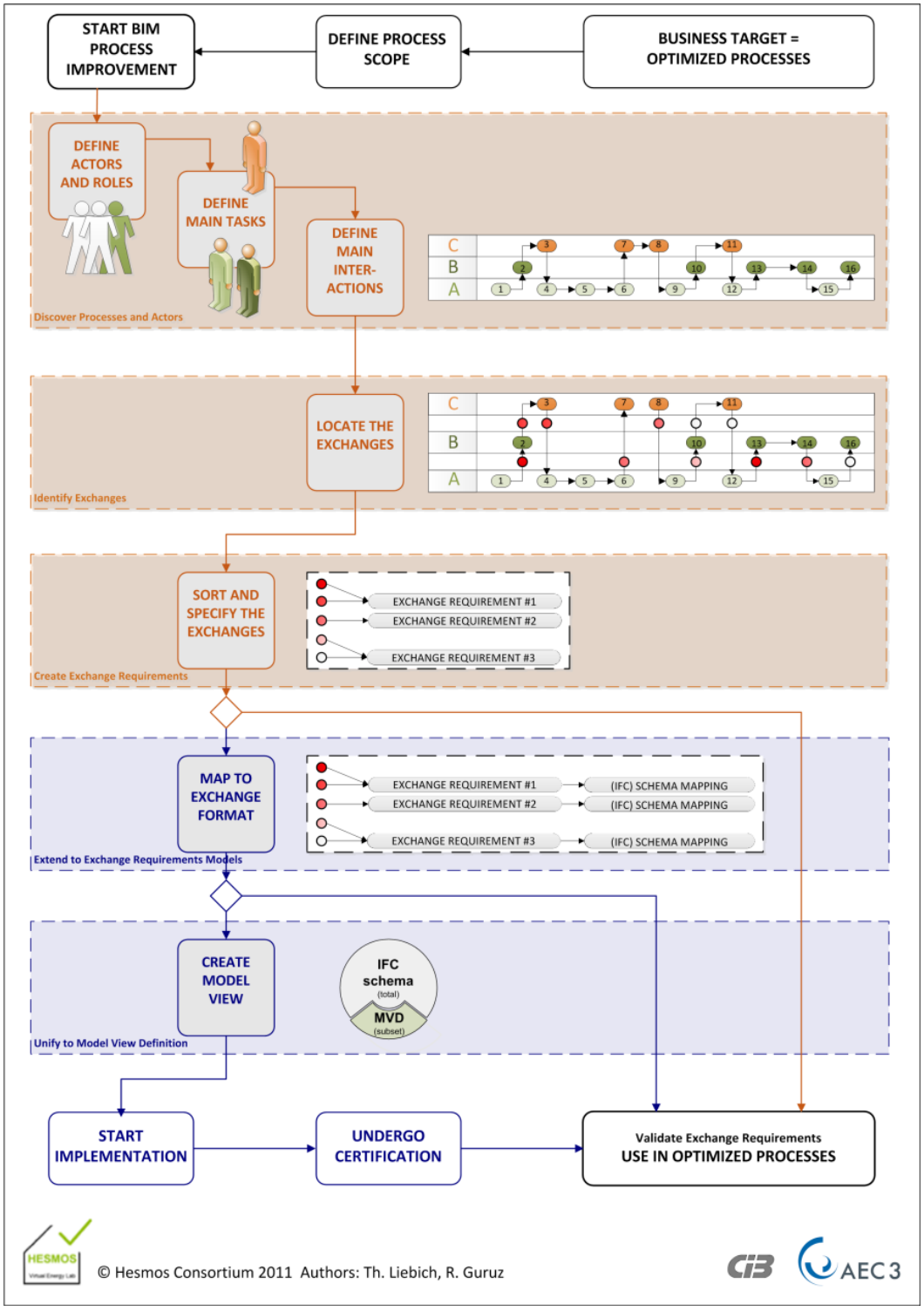

Figure 1 Overview about IDM/MVD (from

3.1.2 Model View Definition

The Model View Definition is translating Exchange Requirements to data structures, which are used for implementation. For the IFC data structure this means to agree on a subset schema of the whole IFC specification and to define additional constraints that needs to be implemented by tool vendors and finally certified by buildingSMART. This not only reduces the efforts for software implementation but will also ensure a certain level of quality for IFC-based data exchange.

MVD developments are not limited to IFC-based data exchange, although existing specification and validation tools may not be used then. In the context of LBD scenarios an MVD could be assigned to one or more (linked) ontologies that are able to cover expected data requirements. This is interesting with respect to data requirements which go beyond BIM/IFC data, either by including other application areas like geographical data (GIS) or by covering a higher level of detail like for instance dealing with special material properties for novel heat loss calculations.

3.2 Adoption in SWIMing

SWIMing is using the IDM/MVD methodology as a reference framework to develop and agree on main criteria for collecting LBD use cases from EeB research projects. These main criteria are:

- stakeholders (actor roles that are involved in tasks)
- building life-cycle stages (high level definitions from feasibility studies to demolition)
- building domains (data exchange definitions using general descriptions)

These criteria enable to cluster and compare use cases on a high level. For those use cases which are identified as having the greatest capability to benefit from adopting BIM-LD technologies, refined versions of the use cases will be developed using BPMN models and more detailed exchange requirements to support the process of converting to LD.

4 Best Practices and Guidelines to Data Management in Relation to SWIMing

The Data Management Plans (DMPs) describes what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. The beneficiaries are expected to take benefits from the generated data in the following manners [3]:

- deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate
- the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;
- other data, including associated metadata, as specified and within the deadlines laid down in the data management plan

The SWIMing project manages the generated data using the following web platforms:

1. Google Drive (private, project internal use only)
   This data is shared within the project consortium are stored and managed in Google Drive. This includes the deliverable drafts, project management documents, presentation slides, project related literatures, etc.
2. SWIMing website (public)
   http://swiming-project.eu/
   The website provides information about the SWIMing project. It informs about the objectives, partners and results of the project. There is also information about all kinds of upcoming events related to topics addressed by SWIMing. Hosted using WordPress, comments can also be added to posts (e.g. events), and made public with the permission of the SWIMing members.
3. W3C LBD Wiki (public)
   https://www.w3.org/community/lbd/wiki/Seed_Use_Cases
   The data related to results of the projects are stored and published in a wikimedia platform. This includes analyzed use cases, data domain categorization, etc. as described in section 2 and the deliverable D1.1. The data is publicly available and editing rights are already granted to registered persons outside the SWIMing project consortium.

Accordingly, data management is not only dealing with public data but also project internal policies. However, the main focus of this Data Management Plan is publicly available data, in particular as SWIMing is actively promoting the reuse of EeB project results.

The SWIMing project not only follows the guidelines on data management in Horizon 2020 as recommended by European Commission [3] but also the best practices of the W3C communities [4, published as draft in June 2015], which SWIMing members are both actively promoting through its dissemination activities in WP3.

The Horizon 2020 guidelines address the following topics:

- *Data set reference and name.* In order to enable identification, search, and retrieval of the data, each data set is named and accessible through a URL. For instance, each business use case identified by SWIMing has its own URL and thus can be referenced as a web resource.
- *Data set description.* Each data set is described by some text including its origin. In the W3C LBD Wiki, it can be seen who are the authors of a certain page. The changes of the pages can be also tracked. Furthermore, in the wiki contents it is also possible to hyperlink to related information or other resources like ontologies or available data sets. These can link to open data silos generated by the project or existing external information sources.
- *Standards and metadata.* SWIMing provides metadata and standardized terms for the W3C LBD Wiki, so that ambiguities and clashes can be avoided. It will give the consumer a better understanding on the collected and enriched data. These terms also act a matrix to compare use cases developed in different projects. The provided data follows the IDM/MVD framework developed by buildingSMART as an open standard for BIM-based use case developments (see section 3). Further details about collected information is provided in deliverable D1.1.
- *Data sharing.* It describes how the data are shared, including access procedures, license, and the management of sensitive data. It will be further explained in section 4.2, 4.4, and 4.7
- *Archiving and preservation.* It deals with the procedures for long-term preservation of the data. It comprises 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. The implementation in SWIMing project is described in section 4.4.

Since SWIMing project uses a web platform to store and manage the generated data and in particular is promoting the use of BLD, it has taken into consideration the best practices for managing data on the web as recommended by W3C. The best practices cover the Data Management Guidelines issued by European Commission. The implementation of each best practice is explained in the following sections.

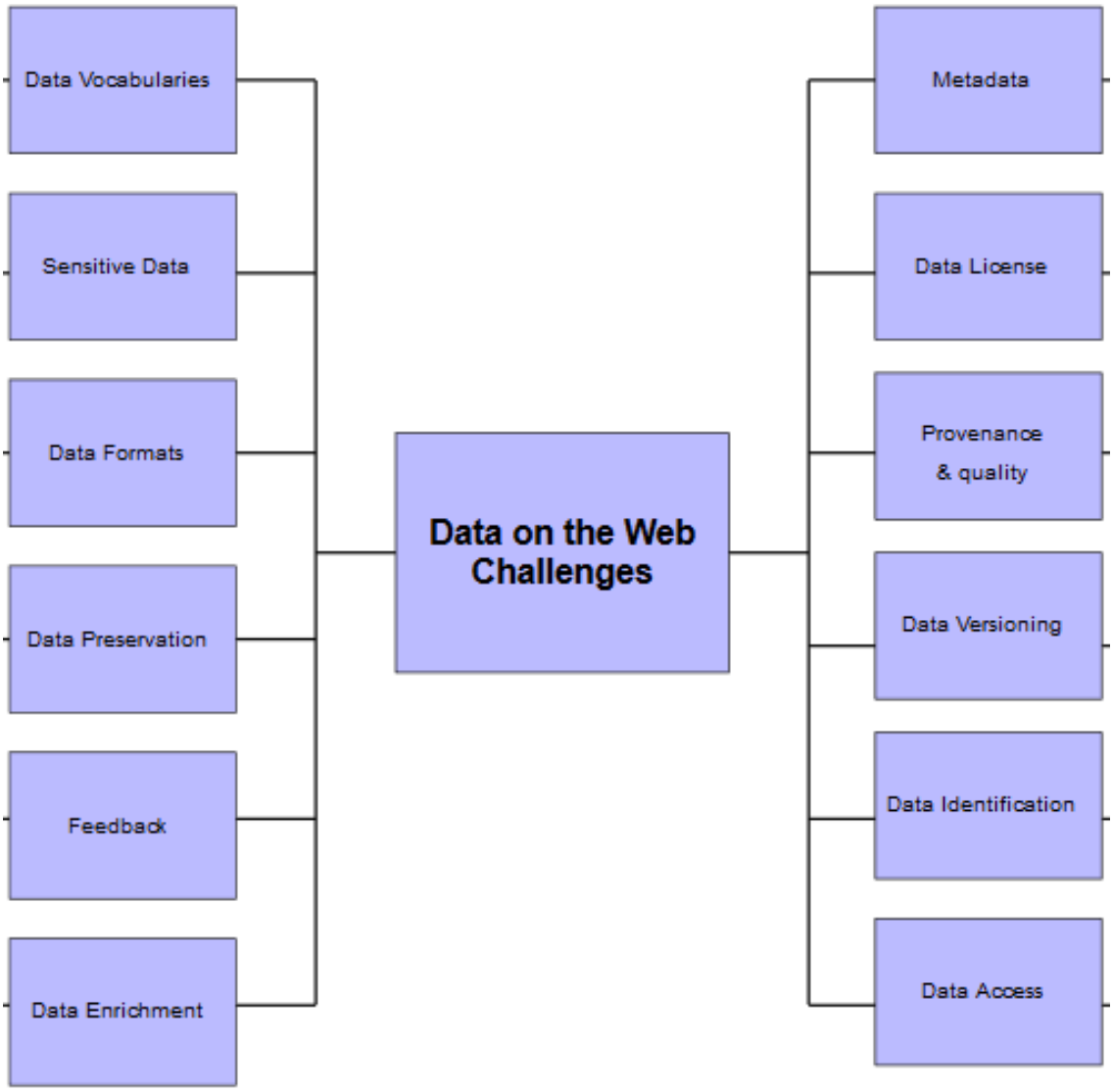

Figure 2 Best practices addressing the challenges faced when publishing data on the Web [4]

4.1 Data Vocabularies and Metadata

This challenge is relevant to achieve semantic interoperability between data producers and consumers. The solution proposed in the Semantic Web community is to agree on a shared vocabulary and to make it available in an open format.

In the W3C LBD Wiki a high level data vocabulary has been developed to share a common understanding about collected data, not only among partners within the SWIMing consortium but also among LBD community members and other external data consumers. It is also intended to avoid ambiguity and clashes as much as possible, this

however remains a challenge due to the wide range and diversity of covered topics. The specific challenge then is to find a good compromise and to keep it as comprehensible as possible.

At the time of this writing the following agreements are made:

1 *Seed Use Case template.* It provides a common structure or template to collect the business use cases. Each of collected use case has to be described and presented in the same way following the template. More information on these can be found on the wiki [2] and also D1.1.

2 *Data domains categorization and taxonomy.* It is an agreed categorization of data domains used by use cases collected from different EU research projects related to energy efficient buildings. Each category is represented by a wiki page, which provides short description, examples of the type of data and some existing RDF- and non-RDF-based data models.

3 *Building Life Cycle Stage.* It lists agreed building life cycle stages considered for analyzing the business use cases, i.e. (i) Planning and Design; (ii) Construction, Commissioning; (iii) Operation; (iv) Retrofitting/ Refurbishment/ Reconfiguration; (v) Demolition/ Recycling.

4 *Stakeholders.* It is an agreed categorization of actors involved in BLC stages such as architect, owner, engineers etc. It includes not only human stakeholders but also organizations like energy supplier or manufacturers and other non-human stakeholders like data providers, applications and devices.

The SWIMing vocabulary has been developed in the beginning of the project and has gone through several steps of refinement. It has been discussed within the LBD community and meanwhile provides a stable basis for our BIM-LD use case collection. However, further refinements and extensions of the vocabulary are very likely to reflect new insights and to deal with requirements coming from use case harmonization and in particular further detailing of key use cases. Extensions and adjustments will be documented on the W3C LBD Wiki to reflect the latest state of the shared vocabulary.

Other agreements have been made for internal work and project management. For instance a simple folder structure based on the work breakdown structure of the work packages is used in our shared Google Docs drive (see Figure 3). Each work package folder contains subfolders corresponding to deliverables. Additional folders are created for other documents like meetings minutes, logos, budget related documents, etc.

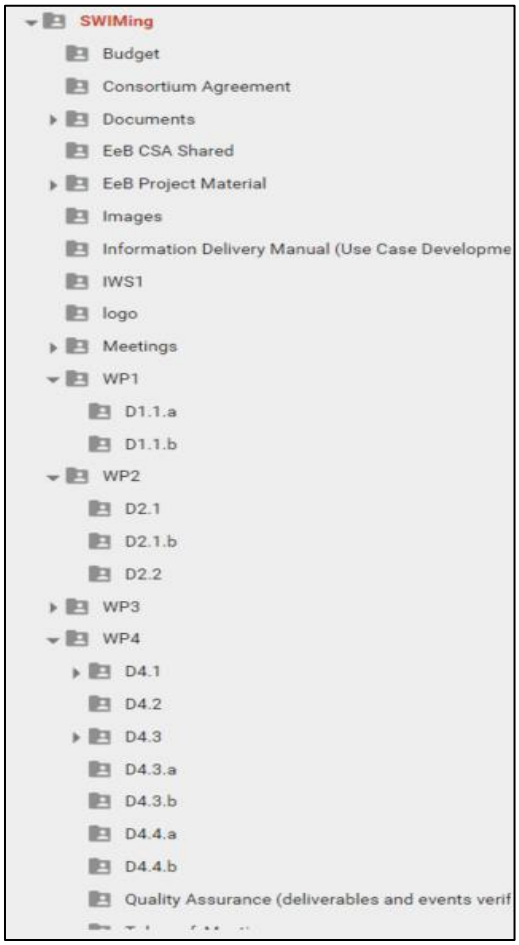

Figure 3 SWIMing Google Drive Folder Structure for internal data management

4.2 Sensitive Data

Sensitive data is any designated data or metadata that is used in limited ways and/or intended for limited audiences. Sensitive data may include personal data, corporate or government data, and mishandling of published sensitive data may lead to damages to individuals or organizations.

To support best practices for publishing sensitive data, data publishers should identify all sensitive data, assess the exposure risk, determine the intended usage, data user audience and any related usage policies, obtain appropriate approval, and determine the appropriate security measures needed to be taken to protect the data. Appropriate security measures should also account for secure authentication and use of HTTPS.

Any use cases generated during the SWIMing project are derived from publicly available deliverables. Where additional data is elicited from the EeB project members, it will only be published on the W3C LBD Wiki with the full permission of the project coordinator. Sensitive data in the form of contacts are only shared through the internal Google Drive and will not be shared without permission of the appropriate party. Data gathered through interviews and questionnaires will also be fully anonymized unless permission is explicitly asked for and given. TCD has its own internal ethics committee which must review any questionnaire or survey before it is used to ensure it complies with its own standards[1] and the standards of the EC[2]. This sets down strict policies for managing and anonymizing personal data.

4.3 Data Formats

Any collected and enriched use case related data is published on a Wiki HTML page that is accessible over the internet. Anyone can access this data, although only members of the community can edit it. So far, main audience of this information are humans as the main aim of this data is to trigger further discussions and information exchange within the LBD community. Accordingly, content of the Wiki pages is mainly structured to meet layout requirements. For further automatic evaluation, especially if collected data is consolidated and amount of data increases, a machine readable format is needed. Ideally, collected data will be offered as RDF graph based on an ontology derived from the vocabulary and agreements discussed in section 4.1. The SWIMing consortium is discussing this option, but has not yet come to a decision.

The internal project data, for example deliverables and project management documents are written using Microsoft Office tools (Word, Excel, PowerPoint). They are also exported in Google Format (Google Docs, Sheets, Slides), so that everyone in the project consortium are able to read and edit the data online. Some of supporting data are represented in PDF format. All used data formats have been selected to optimize data exchange and collaboration within the SWIMing consortium. It is mainly driven by used tools and workflows to reduce coordination overhead.

4.4 Data Preservation

This section describes best practices related to data preservation:

- *The coverage of a dataset should be assessed prior to its preservation* - check whether all the resources used are either already preserved somewhere or provided along with the new dataset considered for preservation.

[1] https://www.tcd.ie/slscs/research/ethics/
[2] http://ec.europa.eu/research/participants/data/ref/fp7/89888/ethics-for-researchers_en.pdf

- *Data depositors willing to send a data dump for long term preservation must use a well-established serialization* - Web data is an abstract data model that can be expressed in different ways (RDF, JSON-LD, ...). Using a well-established serialization of this data increases its chances of re-use.
- *Preserved datasets should be linked with their "live" counterparts* - A link is maintained between the URI of a resource, the most up-to-date description available for it, and preserved descriptions. If the resource does not exist anymore the description should say so and refer to the last preserved description that was available.

All SWIMing data is to be stored on the wiki. Currently, there are no plans to provide the data in other serializations than those provided by the wiki page.

4.5 Feedback

The wiki page is open for the community to contribute and give feedback, SWIMing project members are specifically asking for feedback regarding all EeB project data (relevant to them) published on the wiki and any recommendation to adjust or change that data will be added to the W3C LBD wiki page as received. Feedback is also being elicited through the use of questionnaires and surveys. These are generated using Google forms, which can then be sent to relevant parties. This data is stored on the shared internal Google drive as Google spreadsheets. Paper questionnaires and surveys have also been distributed at workshops and events. This data is also entered into the same google spreadsheets. All feedback during workshops and tutorials will be documented in meeting minutes (e.g. word or google doc) and stored on the shared internal google drive, where they are analyzed by the Steering board and then published on the wiki.

4.6 Data Enrichment

Data enrichment is defined as a set of processes that can be used to enhance, refine or otherwise improve raw or previously processed data [4]. In the SWIMing project, original project documents (deliverables, websites, and specifications) provide the necessary input to extract, categorize and publish required use case related data. This is mainly a review process that requires to harmonize information and, if not available, to enrich data by getting feedback from project partners. References to used resources are always provided so that the original source of information can be used for verification. The review process also includes an assessment regarding the use of BIM-LD (benefits and challenges), which is mainly done by the reviewer as this information shall show the potential as seen by an LBD expert. Other than this, there are no plans for additional enrichment of the data sources generated within the project.

4.7 Data License

A license is a legal document giving official permission to use the data generated or used in a project. According to the type of license adopted by the publisher, there might be more or fewer restrictions on sharing and reusing data. In the context of data on the Web, the license of a dataset can be specified within the data, or outside of it, in a separate document to which it is linked. The SWIMing project will use open web based data and will fully comply with any licenses associated with the data.

4.8 Provenance and quality

Data provenance allows data providers to pass information about the data origin and history to data consumers. It is important to provide it, if the data is shared between collaborators who might not have direct contact to each other, so that the data consumers know the origin or history of the data [4]. In the SWIMing project, the contact data of the author and link to the project homepage, i.e. where the use case originated from, are provided in the use case wiki page. It allows the data consumers to access the original information sources from project home pages and to contact the use case author if necessary.

Furthermore, the wiki platform offers a mechanism to track the changes of each page. The data consumer can see who made the changes and when were the changes made. The changes tracking function is depicted in Figure 4. Whilst the W3C recommends the use of ontologies, e.g. the prov-o ontology[3], to address the challenge in data provenance, the current method of adding and changing use cases on the wiki does not lend itself well to the application of the prov-o ontology. As key use cases are identified and explored in greater detail during the project, the recording of provenance through the use of the prov-o ontology may be applied to support machine readability (see also section 4.3).

Data quality affects the suitability of data for specific applications, including applications. Documenting data quality significantly eases the process of datasets selection, increasing the chances of re-use. Independently from domain-specific peculiarities, the quality of data should be documented and known quality issues should be explicitly stated in metadata [4]. In the project SWIMing the data quality is ensured by asking for feedback from authors/project owners. This will be directly visible in the author's field of collected use cases.

[3] http://www.w3.org/TR/prov-o/

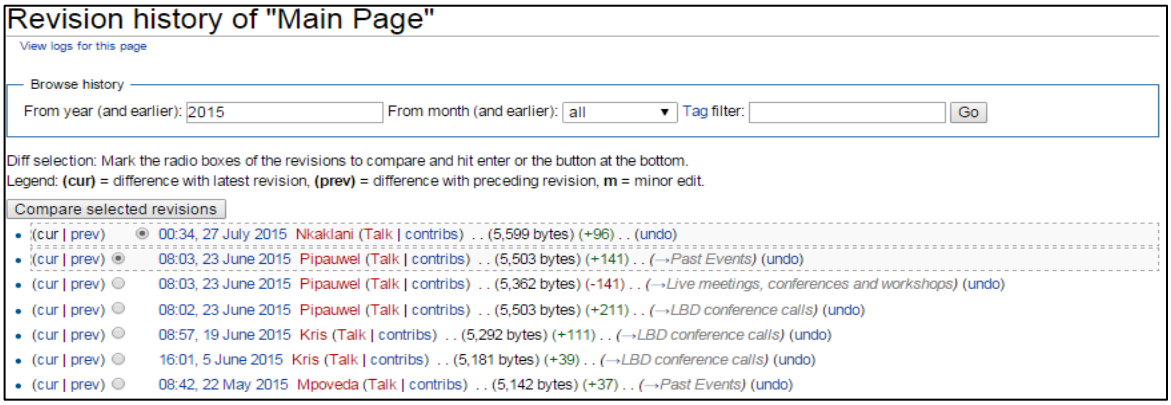

Figure 4 Changes tracking of W3C LBD Wiki

4.9 Data versioning

Data on the web collaboration platform, such as WIki, changes over time. Version information makes a dataset uniquely identifiable. It makes the data consumer to understand how data has changed over time and to determine which version of a dataset they are working with. Good data versioning enables consumers to understand if a newer version of a dataset is available. Explicit versioning allows for repeatability in research, enables comparisons, and prevents confusion [4]. The W3C LBD wiki provides the change log of each wiki page. It can be seen who have performed the changes, when the changes occurred and what exactly the changes are (see Figure 4). Also, the project deliverables will record specific snapshots of the wiki at different times, and these can be further used to track different 'versions' of the use case and data domain classifications and descriptions.

4.10 Data identification

The use of a common identification system helps the data consumers to identify the data and to perform comparison on data in a reliable way. The data has to be discoverable and citable through time. In the SWIMing project, by using the wiki platform, each page containing information about a use case, a data domain category, or a building life cycle stage is accessible through URL. The URL represents the identifier of the corresponding data. It shall not be changed over time. The following gives some example of URL corresponding to use case, data domain category, and building life cycle stage.

- https://www.w3.org/community/lbd/wiki/Building_Energy_Management_System_for_Energy_Efficient_Operation
- https://www.w3.org/community/lbd/wiki/Category:Building_Devices
- https://www.w3.org/community/lbd/wiki/Category:Operation

4.11 Data access

Data consumers usually require a simple and near real time access to data on the web. The W3C LBD Wiki and the SWIMing project website is accessible from anywhere from web browser without any read protection. The SWIMing Google Drive is also accessible from web browser, but only by partners within the consortium. No bulk download neither special APIs is provided for accessing the data other than through HTTP.

4.12 Conclusion of Best Practices and Guidelines

The previous section introduced the best practices guidelines on data management in Horizon 2020 as recommended by European Commission [3] and also the best practices of the W3C communities [4]. It addressed these with respect to the types of data generated by the SWIMing project. This consists of business use cases, in particular those which can benefit from BIM-LD, and also guidelines and best practices for converting building data to LD. This data will be stored on the shared W3C portal and wiki and as such, we do not at this stage foresee the need for ontological descriptions of the data, in particular, for recording provenance, licensing etc. The types of data that projects which SWIMing is clustering though will benefit from these same guidelines, and so, the project will be actively promoting their usage during events held as part of WP3 dissemination and clustering. In the next section we examine how SWIMing compliments the CSA Ready4SmartCities, which has looked also at the application of LD technologies in the Smart City domain.

5 Comparison with Ready4SmartCities

The Ready4SmartCities project presented a set of guidelines for Linked Data generation in the energy domain [5] aiming to address:

- The generation of Linked Data from tabular (SQL, XLS, or CSV) file formats, among others, which are the formats that are currently the most used in the energy domain.
- The issue of legal aspects, licenses, and data ownership, which is regarded as an important topic that could help lowering the barrier to publish data.
- The generation of static data, as well as dynamic data.
- Various means of obtaining and accessing the data, including data stored in files, which is in line with the specified requirements.

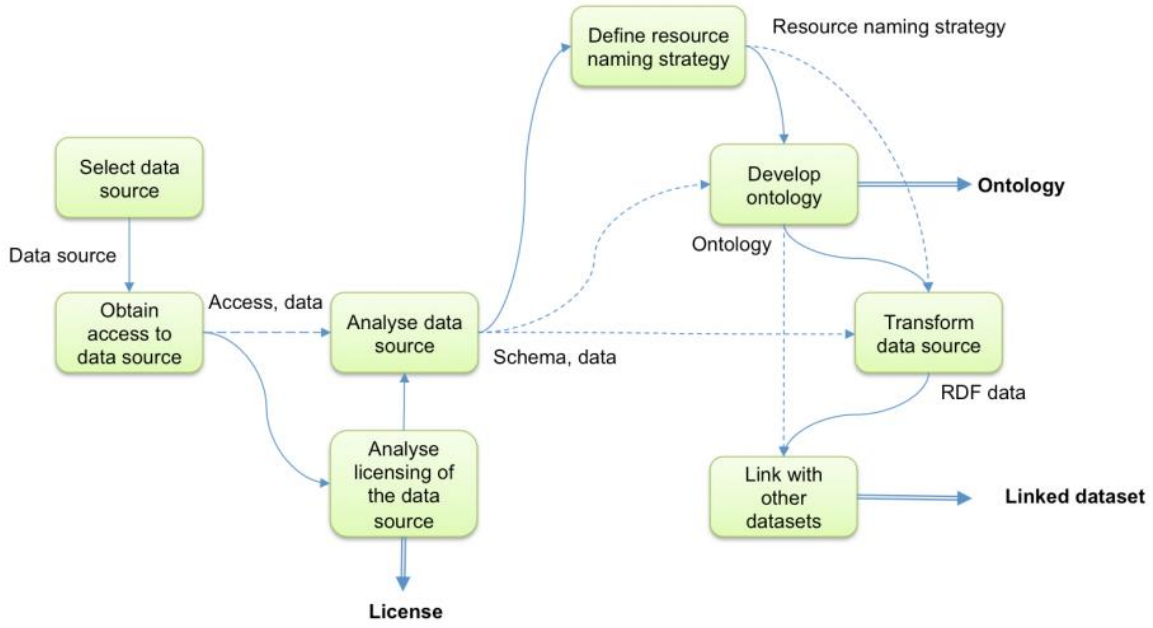

Figure 5 Ready4SmartCities - Steps of the guidelines for Linked Data generation [5]

Figure 5 presents the generic steps for generating Linked Data as proposed by Ready4SmartCities. Moreover, a set of requirements for the publication of Linked Data in the energy domain has been also introduced by Ready4SmartCities [6] provided in a consolidated way together with two available standards:

- the ISO/IEC 25012 standard (International Organization for Standardization) on Data Quality for the scope of Linked Open Data that provides some data quality indicators which are analyzed for quality requirements extraction, and

- the AENOR (La Asociación Española de Normalización y Certificación) PNE 178301 Spanish standard on Smart Cities and Open Data, which presents a set of metrics and indicators concerning the maturity of the opening and publishing data from the public sector in order to facilitate their reuse for the scope of Smart Cities.

The overall requirements extracted by the research and survey analyses are summarized into the categories presented in Figure 6. READY4SmartCities aimed at identifying existing knowledge and data resources that are independent from the Energy Management Systems (EMS) domain, as well as ontologies, datasets and alignments specific for EMS interoperability [7]. For the collection of ontologies and datasets, a special online catalogue [8] has been developed to ensure that resources are collected and recorded in a standardized way.

The catalogue also allows for ease of understanding and use in terms of submission of new content, visualization of existing resources and handling of recorded items. For the collection of alignments, an alignment server offered as a web service has been set up in order to identify and document links and alignments among the identified ontologies and datasets.

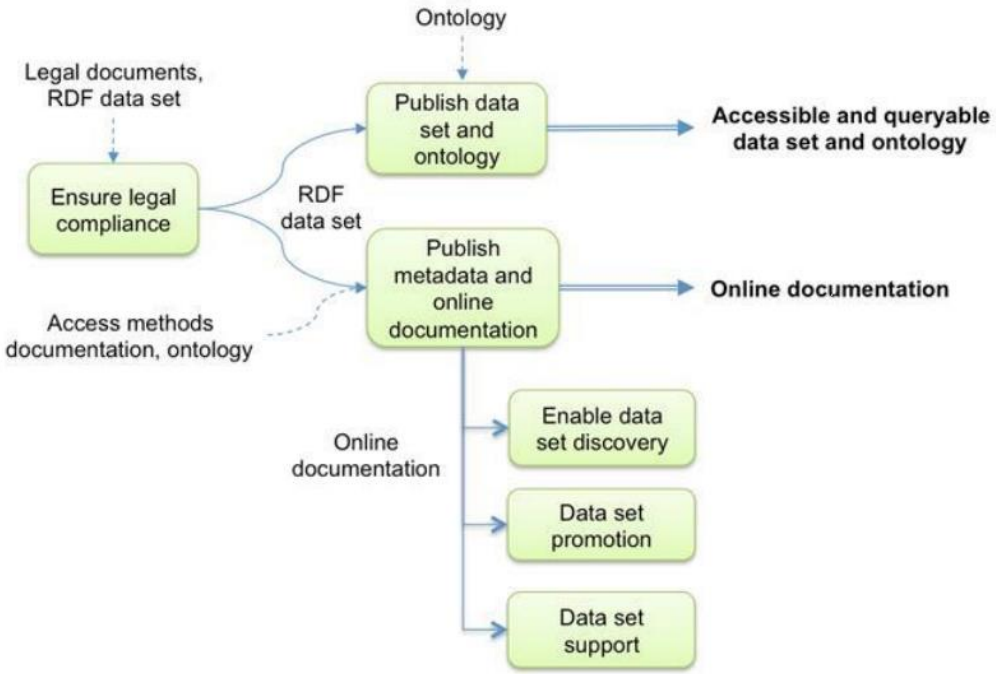

Figure 6 Ready4SmartCities - Tasks for Linked Data publication [6]

While READY4SmartCities is mainly focused on identifying energy-related ontologies and datasets, the SWIMing project has a complementary scope by identifying and analyzing business use cases for Building Information Modelling (BIM) and Linked Data. SWIMing further analyses their potential extensions to better represent issues such as data modality and data format, with the goal of enabling fully automatic discovery and consumption of resources by Building Life Cycle Energy Management (BLCEM) systems.

6 Conclusion

The Data Management Plan (DMP), which is a requirement for all projects participating in the H2020 Pilot on Open Research Data, aims to maximize access to and re-use of the research data generated during the course of the project. The SWIMing project is a Coordination & Support Action and does not actively research on topics related to Energy-efficient Buildings (EeB) and the use of Linked Building Data (LBD). The aim of SWIMing is rather to extract and share knowledge generated by various EeB projects. The main source of data generated by SWIMing is the LBD wiki, which provides a portal for the community to access and contribute toward descriptions of business use cases. Data will also be generated in the form of guidelines and best practices for generating free interlinked, and semantically interoperable BIM resources for meeting current and future application requirements within the BLC

For the structuring of use cases, this deliverable documents a standard methodology to capture those use cases (IDM/MVD) and provides a description on how the generated data is to be stored, made accessible for verification and re-use, and how it is being curated and preserved via the shared community W3C community portal. The document also presents the best practices as set down by the W3C on publishing data on the web and R4SC project, which both address the same concerns as the DMP guidelines. As a CSA, SWIMing will be actively promoting these best practices amongst the wider EeB communities, and will be providing the expertise and tools to those projects who are unfamiliar with these practices so that they may apply them to their own project generated data, thus supporting greater exploitation of their project results and thus increase impact for their project outcomes.

References

[1] R. Liebich, T., Stuhlmacher, K., Weise, M., Katranuschkov, P. & Guruz, "HESMOS Deliverable D2.1 BIM Enhancement Specification."

[2] "W3C LBD Wiki- Seed Use Cases." [Online]. Available: https://www.w3.org/community/lbd/wiki/Seed_Use_Cases.

[3] "European Commission, Guidelines on Data Management in Horizon 2020, Version 1.0." [Online]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

[4] "W3C, Data on the Web Best Practices, Working Draft." [Online]. Available: http://www.w3.org/TR/2015/WD-dwbp-20150625/.

[5] "Ready4SmartCities D4.1: Requirements and guidelines for energy data generation."

[6] "Ready4SmartCities D4.2 - Requirements and guidelines for energy data publication."

[7] "Ready4SmartCities D2.2: Ontologies and datasets for Energy Management System interoperability v1."

[8] "Ready4SmartCities Online Catalogue of Ontologies." [Online]. Available: http://smartcity.linkeddata.es/. [Accessed: 27-Jul-2015].

[9] ISO 29481-1:2010 "Building information modelling -- Information delivery manual -- Part 1: Methodology and format".