D8.5 Market Study and Exploitation Plan

DURAARK

FP7 – ICT – Digital Preservation
Grant agreement No.: 600908

Date: 2015-03-18
Version 1.1
Document id. : duraark/2015/D.8.5/v1.1

Grant agreement number	:	600908
Project acronym	:	DURAARK
Project full title	:	Durable Architectural Knowledge
Project website	:	www.duraark.eu
Partners	:	LUH – Gottfried Wilhelm Leibniz Universitaet Hannover (Coordinator) [DE] UBO – Rheinische Friedrich-Wilhelms-Universitaet Bonn [DE] FhA – Fraunhofer Austria Research GmbH [AT] TUE – Technische Universiteit Eindhoven [NL] CITA – Kunstakademiets Arkitektskole [DK] LTU – Luleå Tekniska Universitet [SE] Catenda – Catenda AS [NO]
Project instrument	:	EU FP7 Collaborative Project
Project thematic priority	:	Information and Communication Technologies (ICT) Digital Preservation
Project start date	:	2013-02-01
Project duration	:	36 months
Document number	:	duraark/2015/D.8.5/v1.1
Title of document	:	D8.5 Market Study and Exploitation Plan v1
Deliverable type	:	Report
Contractual date of delivery	:	2015-01-31
Actual date of delivery	:	2015-03-18
Lead beneficiary	:	LTU
Author(s)	:	Dag Fjeld Edvardsen `<dag.fjeld.edvardsen@catenda.no>` (Catenda), Jakob Beetz `<j.beetz@tue.nl>` (TUE), Martin Hecher `<martin.hecher@vc.fraunhofer.at>` (FhA), Martin Tamke `<martin.tamke@kadk.dk>` (CITA), Michelle Lindlar `<michelle.lindlar@tib.uni-hannover.de>` (LUH), Richard Vock `<vock@cs.uni-bonn.de>` (UBO), Sebastian Ochmann `<ochmann@cs.uni-bonn.de>` (UBO), Stefan Dietze `<dietze@L3S.de>` (LUH), Frode Randers `<frode.randers@ltu.se>` (LTU)
Responsible editor(s)	:	Frode Randers `<frode.randers@ltu.se>` (LTU)
Quality assessor(s)	:	Eva Eggeling `<eva.eggeling@fraunhofer.at>` (FhA) Raoul Wessel `<wesselr@cs.uni-bonn.de>` (UBO)
Approval of this deliverable	:	Stefan Dietze `<dietze@L3S.de>` (LUH) – Project Coordinator
Distribution	:	Public
Keywords list	:	market study, exploitation plan, sustainability plan

Executive summary

This deliverable (D8.5 Market Study and Exploitation Plan V1) reports on plans for the continued, sustainable use of the research results achieved in the DURAARK project. Identified gaps in current markets and potentials for a commercialisation are introduced and discussed.

The focus of this document lies on the exploitation activities that are carried on during the project and pave the road for subsequent activities after successful completion of the project. It will be updated by another version (D8.7) which will focus on exploitation activities after the project has ended.

Table of Contents

1 Introduction

This report describes a first version of the DURAARK consortium's exploitation plan to achieve sustainable use of the project's outcomes. The foundation for the report is based on

- discussions with the stakeholders
- formal and informal interviews
- market, gap and requirement analyses
- feedback received from the advisory board and members of the various research communities that the preliminary results have been introduced to

as well as other assessment methods. Insights from earlier deliverables such as D2.2.1 "Requirements Document", D7.7.1 "Current state of 3D object processing in architectural research and practice", D6.6.1 "Current state of 3D object digital preservation and gap-analysis report" and their comments have been incorporated in the strategies and plans laid out in this document.

The main stakeholder groups in the DURAARK project are (as described in deliverables D2.2.1 and D7.3):

- Architects and engineers
- Construction companies
- Researchers and lawyers
- Building owners and real estate managers
- Public administrations/ Public planning / Policy makers
- Knowledge base maintainers
- Cultural heritage institutions
- Suppliers of software and IT services to facility managers
- Research communities in areas such as digital preservation, building information modelling and semantic web/linked data

These groups of stakeholders and their specific requirements and characteristics with respect to exploitation are further described in Section 3. A detailed description different groups as well as use-cases can also be found in section 3 of the requirements document D2.2.1

In addition to earlier reports analysing the stakeholders and use cases, complementary documents for this deliverable are:

- Deliverable D7.3 contains the analysis of the workshop arranged 12th of November 2014 in Copenhagen with representatives from the stakeholder groups and, as such, gives a good background to understanding the stakeholders' needs.
- Deliverable D1.7 describes the IPR Management plan for the project, in particular detailing how licensing and IPR management supports exploitation of DURAARK outcomes.
- Deliverable D8.6 reports dissemination activities, which support and complement the exploitation activities described here.

Information that is provided in these other deliverables, are presented here in a condensed form in order to keep this document informative and reduce redundancy. Where appropriate, we refer to the original deliverable for a more elaborate description.

2 Exploitable outcomes

A very detailed description of tangible IP and exploitable outcomes from the DURAARK project is listed in the IPR Management Plan (deliverable D1.7). Here, we provide an overview and description of these items, particularly composed of knowledge and best practices (in the form of reports and publications), data and software.

2.1 Knowledge and best practices: reports and publications

This category includes all IP covered in written reports and documents, including:

- scholarly publications[1],
- technical reports, and
- deliverables[2], most notably deliverable of type R (report).

Reports typically provide the theoretical background and documented knowledge related to other outcomes, such as software and data, and ensure the reproducibility of project results. As such, they provide and convey the knowledge gained throughout the project and the expertise required to use, reproduce, exploit or implement techniques, data or software produced in the technical work packages of DURAARK.

2.2 Software artefacts and services

The software artefacts created in DURAARK play an important role in attracting practitioners in the field, providing examples of functionality and acting as a starting point for discussions with the (commercial) stakeholders. The software artefacts created in DURAARK are structured into three categories:

Integrated Software Prototype The integrated software prototype is a **graphical user interface** for stakeholders to access the functionalities developed in DURAARK in an easy-to-use, predefined way. The project name for the prototype is **WorkbenchUI**.

Service Platform DURAARK comprises areas like , e.g. semantic enrichment, long-term archival, etc. (see below for a full list). All the functionalities provided

[1] http://duraark.eu/publications/
[2] http://duraark.eu/deliverables/

in those areas are delivered to a stakeholder (or "consumer", in technical terms) via **services**. The conglomerate of thematic services developed in DURAARK is called the "Service Platform". This service-oriented architecture allows a) easy **integration** of DURAARK's functionality into other projects and applications and b) easy **extension** of the platform by third parties in adding custom services.

Python-Bindings to Geometric Enrichment Services The Python-bindings for accessing parts of the geometric enrichment services (described in Section 2.2.2.2) is a complement to the Service Platform regarding the integration of functionality into commercial (and also non-commercial) products in the field of architecture. We are providing a Python interface to the tools for easy integration into products for the Architectural and Engineering Construction Industry (AEC). Python has become a widely used programming interface for most AEC related programs and is used by Software distributors, Architects and Engineers to develop business and project specific solutions. Environments, where parts of the Geometric Enrichment Services been successfully implemented are McNeel Rhinoceros[3], which supports natively Python[4], the related and highly popular parametric visual scripting environment Grasshopper[5], AutoDesk Revit[6] and its Pythonshell[7], the related and quickly evolving parametric visual scripting environment Dynamo[8]. Other AEC software with Python interfaces are for instance ArchiCAD[9] and CATIA[10].

The following sections describe how the developed functionality is provided and how it can be exploited or consumed by stakeholders. Additionally, it depicts the involved software components. Section 2.2.1 targets stakeholders with an end-user's point of view, section 2.2.2 focuses on stakeholders who want to integrate or extend DURAARK's functionality via services. Section 2.2.3 shows which outcomes support the integration of DURAARK's functionality in commercial and non-commercial products in the field of architecture.

[3] https://www.rhino3d.com/
[4] http://wiki.mcneel.com/developer/python
[5] http://www.grasshopper3d.com/
[6] http://www.autodesk.com/products/revit-family/overview
[7] https://code.google.com/p/revitpythonshell/
[8] http://dynamobim.com/
[9] http://www.graphisoft.com/archicad/
[10] http://www.3ds.com/products-services/catia/

2.2.1 WorkbenchUI

The WorkbenchUI is a software component that provides a **graphical user interface** (GUI) to the functionality developed in DURAARK. Targeted stakeholders are **end-users** who want to use an application to perform long-term archival tasks in the field of architectural data. The WorkbenchUI compiles experience and best practices into **pre-defined procedures** in a wizard style that also reflects best practises in existing archival tools. The graphical user interface guides the end-user in a **step-by-step** manner through those procedures. In this sense the WorkbenchUI is a **reference implementation** of the knowledge gained in the project, including the following areas (derived from the use case definitions in deliverable D1.7):

- Semantic enrichment of architectural content (responsible: WP3)
- Geometric enrichment of architectural content (responsible: WP4/5)
- Preservation of architectural content (responsible: WP6)
- Search & Retrieval of archived architectural content (responsible: WP2)

The WorkbenchUI as graphical user interface relies on **services** and **consumes** functionality provided by those services. For instance: An end-user has a point-cloud file and wants to reconstruct an IFC file from it. Via the WorkbenchUI s(he) selects a) the file and b) the IFC reconstruction operation in the GUI. The WorkbenchUI then delegates the an IFC reconstruction assignment to the geometric enrichment service, which provides this operation. After the service finishes, the GUI is notified and displays the result. All operations provided by the WorkbenchUI are following this workflow. The services which are performing the actual work are summarised in the **Service Platform**, described in the next Section 2.2.2. The strict separation between graphical user interface and service platform allows to easily develop alternative GUIs for DURAARK's services. This is useful in case the WorkbenchUI does not meet the requirements of stakeholders or other third parties. It also allows the reuse of individual modules in other existing archival frameworks[11].

[11] like the Open Source Archivematica platform

Software artefacts

Name	WorkbenchUI
Type	Web-Application
Demo	`http://workbench.duraark.eu`
Source	`http://github.com/DURAARK/workbench-ui`
Responsibility	WP2

To give the reader a better impression and understanding, Figure 1 and Figure 2 show screenshots of the WorkbenchUI along with a short description.

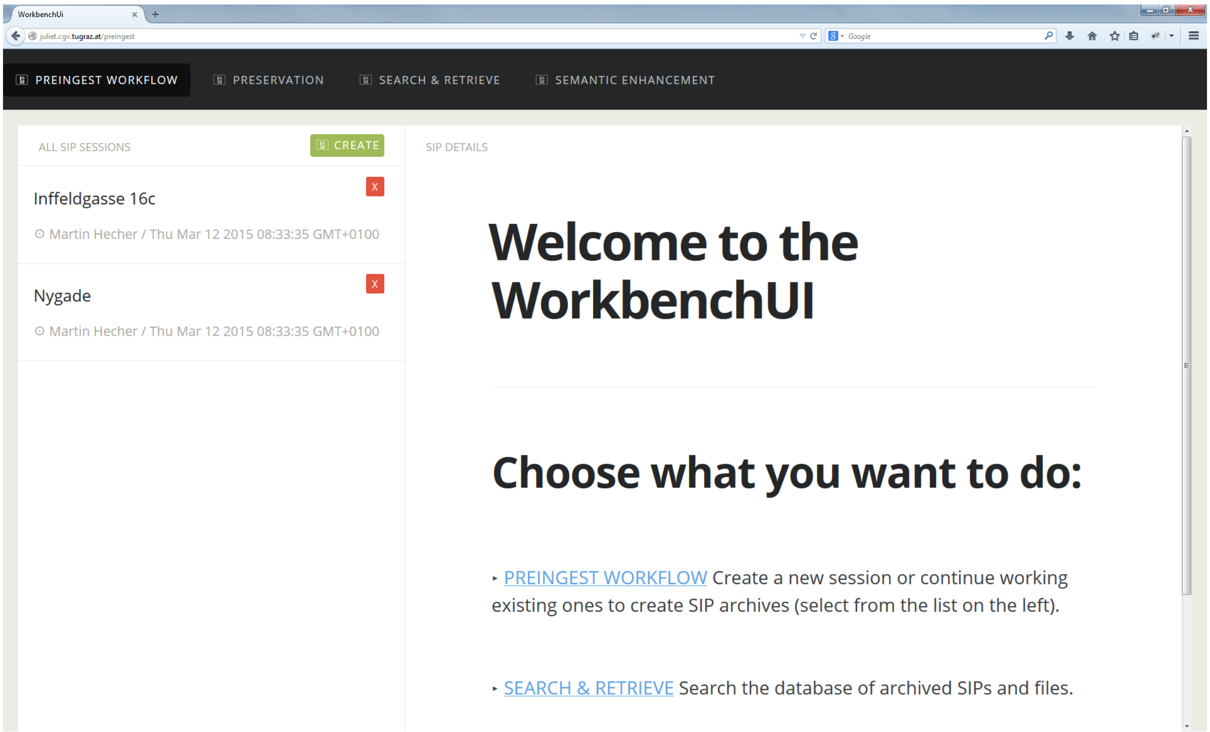


Figure 1: Selection of a session for creating a long-term preservation archive.

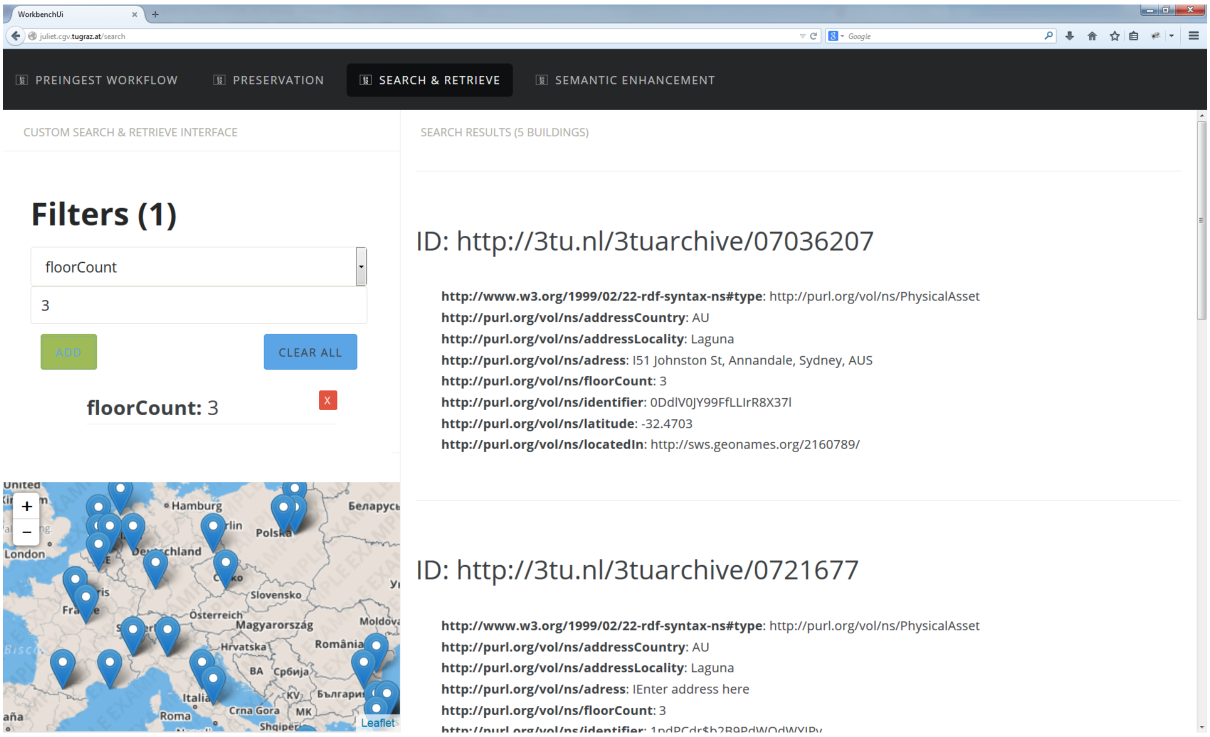


Figure 2: Search and retrieve interface for the DURAARK WorkbenchUI. Partial metadata records stored in the Semantic Digital Archive (SDA) are displayed in the main content area (right hand side). A search can be initiated by using filters (upper left) or selecting geo-referenced assets from the interactive map (lower left).

2.2.2 Service Platform

The Service Platform bundles the functionality developed in DURAARK in an easy-to-use application programming interface (API). The API is implemented using a RESTful[12,13] service-oriented architecture (SOA). At the core of the DURAARK platform, the Service Platform is implemented as a **Web-API**.

In the last decade Web-APIs became increasingly important for developing distributed, service-oriented applications. Figure 3 shows the growth of such APIs starting since

[12] http://en.wikipedia.org/wiki/Representational_state_transfer

[13] The Service Platform is not implementing all principles of original REST description, as the state-of-the-art Web-APIs are also only using parts of the specification for simplicity. The web community is currently discussing on how to best apply more REST principles to current APIs, e.g. in explicitly using HATEOAS (http://en.wikipedia.org/wiki/HATEOAS).

the year 2005. Well-known examples of such Web-APIs, which have an immediate use in the respective products or are re-used in many third-party applications, are Google Maps, Facebook, Paypal, LinkedIn or Youtube. The end-user normally does not work directly with those APIs, but the respective graphical user interfaces are well known. In DURAARK we follow the same pattern for presenting functionality to an eventual stakeholder: the Service Platform is the Web-API backbone of the system that is invisible to the end-user. The WorkbenchUI, as the graphical user interface, guides users through the different workflow processes in the use case scenarios in long-term archival of architectural data.

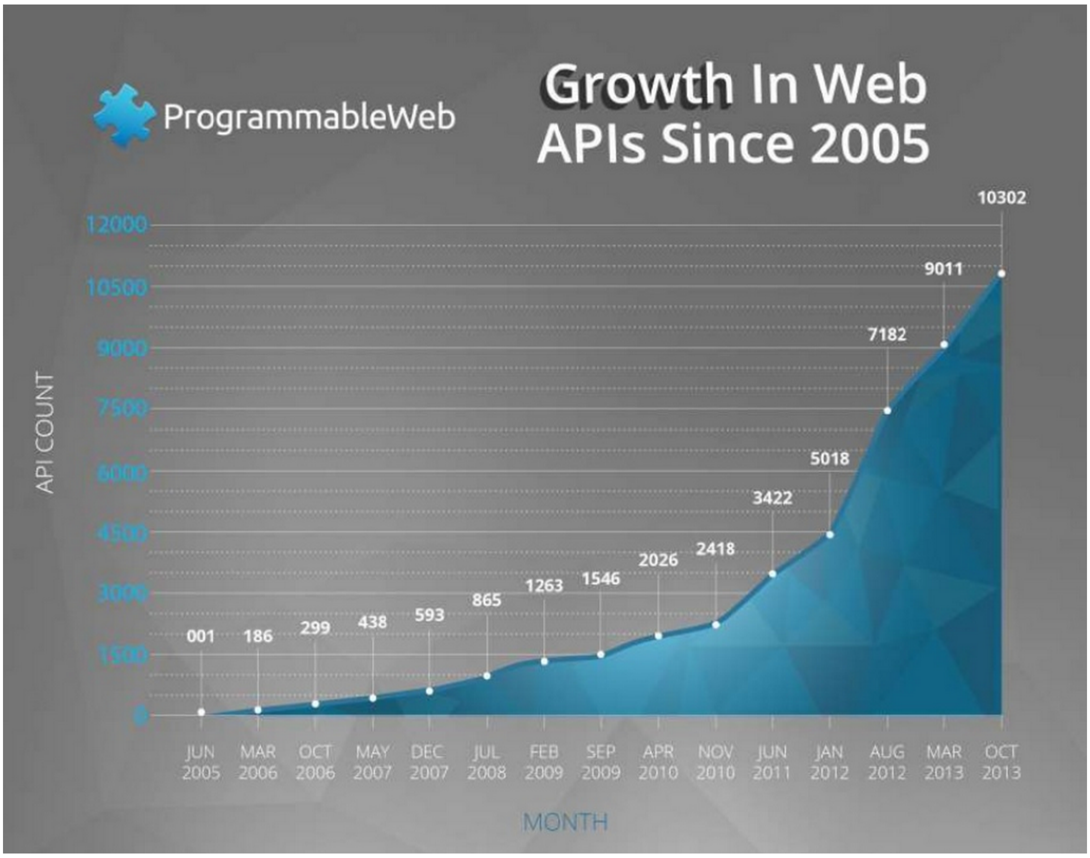


Figure 3: Growth in Web-APIs since 2005 until 2013 (Source: http://www.slideshare.net/programmableweb/web-api-growthsince2005).

Three main advantages of such service-oriented architecture (SOA) approaches can be identified:

Decoupled Logic and Views It is not mandatory to use the WorkbenchUI as the graphical user interface for DURAARK. Depending on the use case, an alternative user interface that communicates with the API and triggers operations can be developed.

Modularity The Service Platform provides functionality modules (see sections below). Via its API it is possible for third party developers to integrate all or parts of the functionality into their own (web-)applications. Examples for such principles are Facebook's "Like-button", the possibility to embed Youtube videos in third-party websites or the integration of PayPal in your web-shop systems. Analogue to these principles of re-usability the Service Platform makes it possible to embed functional pieces developed in DURAARK into other systems that can be either web- or desktop-based.

Extensibility The service-oriented nature of the Service Platform makes it easy to add new services by third parties. As part of the work in WP2, tools are provided to help a developer generating the source code for a new service[14] and use this code-blueprint to add custom functionality.

The Service Platform is structured into the following functional areas and sub-services:

- Semantic enrichment
- Geometric enrichment
- Preservation of data
- Search & Retrieval of archived data

Each sub-service follows the same structure:

1. The functionality itself is provided by a software component developed in one or multiple work packages.
2. The software component is wrapped into a service by adding a REST-API layer on top of the component.

The following sections give a short description of the purpose of each area or sub-service and list the software artefacts produced by the service. The software artefacts are divided into the software components produced and the respective service implementation.

[14]See https://github.com/DURAARK/generator-duraark

2.2.2.1 Semantic Enrichment Service

We refer to the semantic enrichment as a process consisting of several steps dealing with addition of the building information stored in the DURAARK archival platform. The enriched building information is captured as part of the SDA and structured in accordance to the `buildM` metadata schema (see D3.3 and D6.2).

The components that are used in this process are: (i) E57 Metadata extraction and (ii) IFC metadata extraction. The extraction modules operate on the two main data types addressed by the DURAARK system (IFC and E57) and are complemented by additional modules that enrich the extracted metadata with contextual information. An example for these enrichment modules is the dataset crawler module, which performs targeted searches on individual information bits of the metadata (usually consisting of entities of type building). The details about the corresponding components are described in previous deliverables D3.3 and D3.4.

Software artefacts The software artefacts created for the semantic enrichment service are listed in Table 1. The Software components themselves can be reused for e.g. extracting metadata from IFC files outside the DURAARK context.

The screenshots in Figure 4 and Figure 5 show the graphical user interface for the described services where applicable. The GUI in this case is the WorkbenchUI, as the services (or more precisely the underlying software components) do not provide their own GUI.

2.2.2.2 Geometric Enrichment Service

In order to allow efficient long-term archival of 3D BIM models and point-cloud scans and to ensure future searchability and access, several components within the DURAARK framework focus on the geometric representation of the data rather than on associated metadata. To ensure a consistent documentation and archival of buildings, users are offered tools to synchronise and link 3D BIM models and point-cloud scans from different time frames, offering future stakeholder the possibility to visualise the geometric changes the building underwent. Additionally, DURAARK offers components for semi-automated metadata enrichment of point-clouds, e.g. by generating information about room sizes

Name	Type	Wraps	Source code	Lead by
microservice-ifcmetadata	Web-API	pyIfcExtract	microservice-ifcmetadata	WP2
microservice-e57metadata	Web-API	e57Extract	microservice-e57metadata	WP2
microservice-semenrichment	Web-API	ifcContextual-Enrichment	microservice-semenrichment	WP2
microservice-sda	Web-API	SDA	microservice-sda	WP2
pyIfcExtract	Software component	N/A	pyIfcExtract	WP3
e57Extract	Software component	N/A	e57Extract	WP5
ifcContextual-Enrichment	Software component	N/A	ifcContextual-Enrichment	WP3
Semantic Digital Archive (SDA)	Software component	N/A	t.b.d.	WP3

Table 1: Listing of the published software components and respective Web-API wrappers. Source code repositories reside under the same name as the component in the http://github.com/DURAARK/ realm.

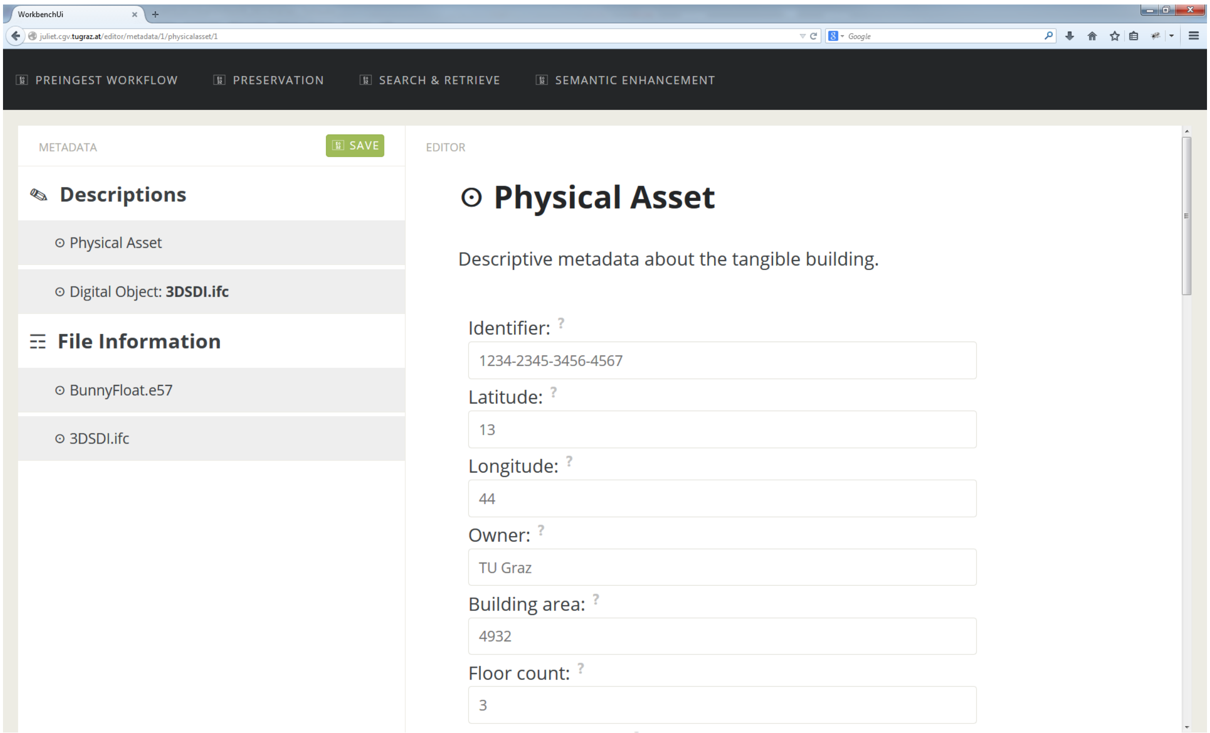


Figure 4: GUI for extracting metadata from given IFC and e57 files and for working with the extracted metadata.

and their connectivity. This allows users the ability to search formerly unstructured point-clouds for which no textual description were given, which greatly increases the usability of the archive.

Another important component provides stakeholders the possibility to create lightweight access copies from 3D point-clouds. Due to modern scanning technologies, point-clouds can easily have sizes up to several hundreds or even thousands of gigabytes, rendering them extremely difficult to handle. On the one hand, DURAARK components are able to generate small IFC models from the point-cloud, which are ideal for preview over a network connection. On the other hand, there are also components that allow compressing the point-cloud by exploiting certain geometric regularities, rendering e.g. delivery over a network more manageable.

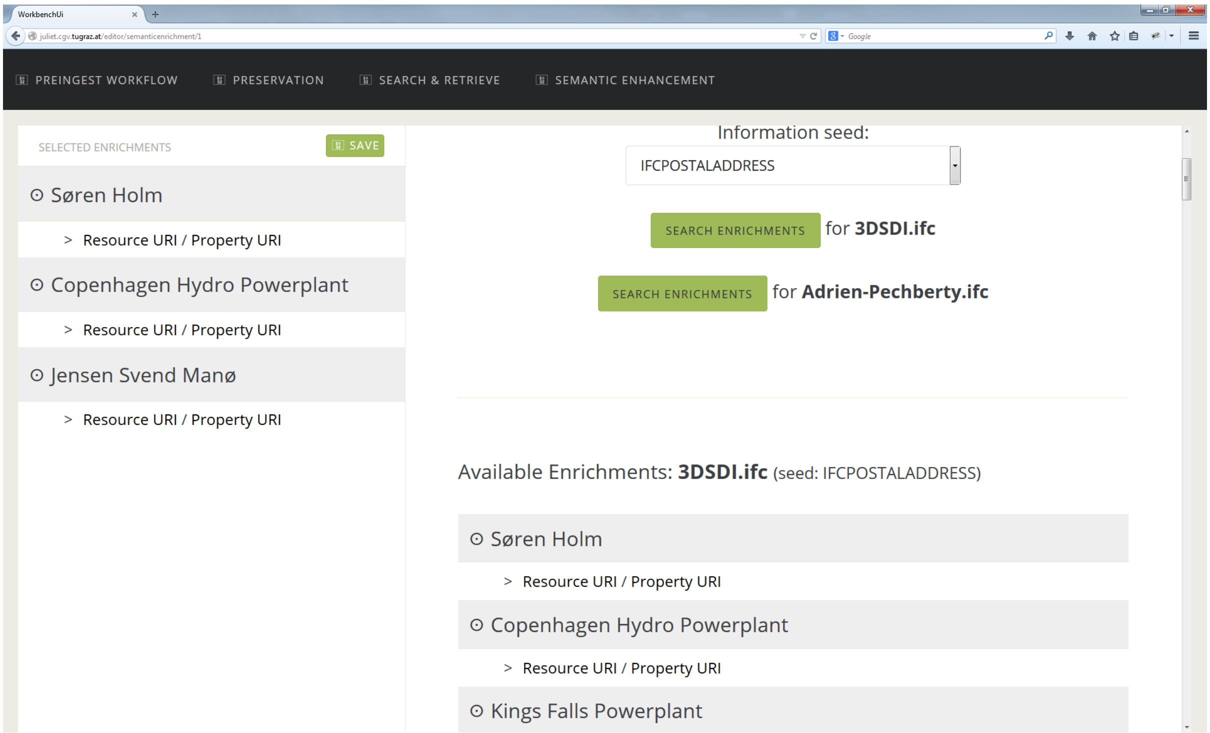


Figure 5: Interface to decide which semantic enrichment (based on the extracted metadata) to include into the archive.

Software artefacts

Name	microservice-differencedetection
Type	Web-API
SW-Component	DifferenceDetection
Source	http://github.com/DURAARK/microservice-differencedetection
Responsibility	WP2

Name	microservice-ifcreconstruction
Type	Web-API
SW-Component	IFCReconstruction
Source	http://github.com/DURAARK/microservice-ifcreconstruction
Responsibility	WP2

Name	microservice-rise
Type	Web-API
SW-Component	RISE
Source	http://github.com/DURAARK/microservice-rise
Responsibility	WP2

Name	DifferenceDetection
Type	Software component
Python-API	Yes (see 2.2.3)
Source	Closed source license
Responsibility	WP4

Name	IFCReconstruction
Type	Software component
Python-API	No
Source	TBA
Responsibility	WP5

Name	RISE
Type	Software component
Python-API	No
Source	http://github.com/DURAARK/RISE
Responsibility	WP5

The screenshots in Figure 6, Figure 7 and Figure 8 show the graphical user interfaces for the described services or output of the services.

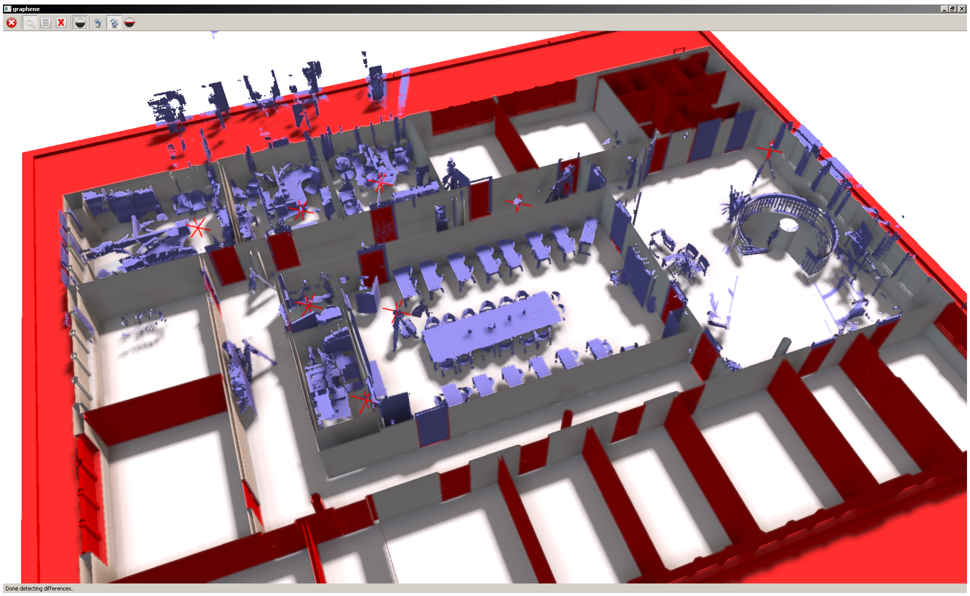


Figure 6: Standalone graphical user interface for the DifferenceDetection component. The WorkbenchUI is using the functionality via its service implementation.

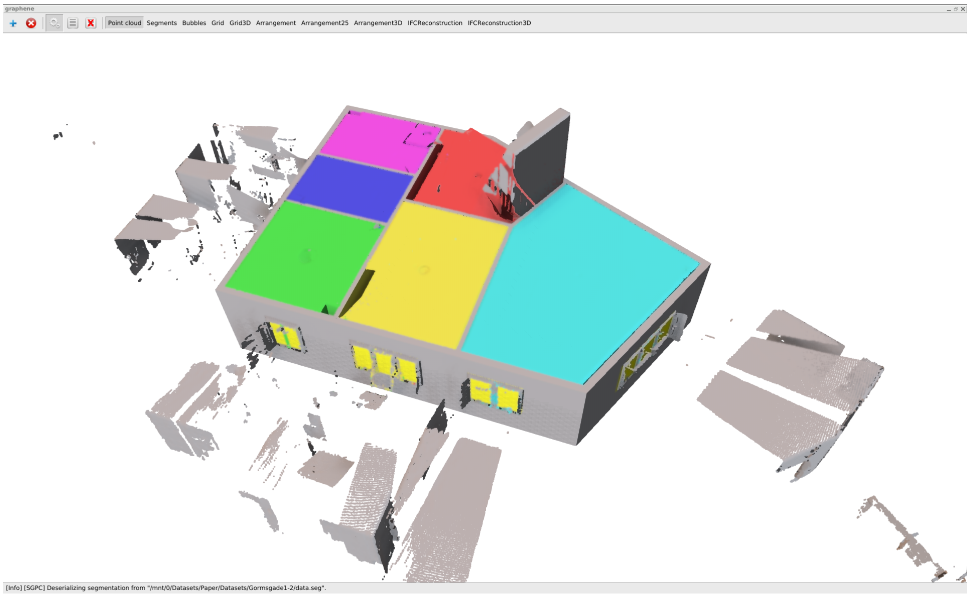
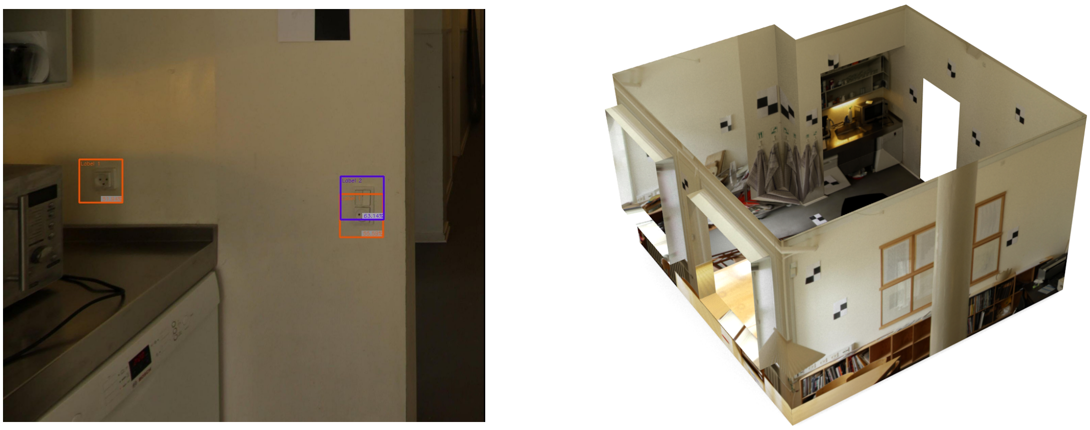


Figure 7: Standalone graphical user interface for the IFCReconstruction component. The WorkbenchUI is using the functionality via its service implementation.



(a) (b)

Figure 8: Output of the RISE software component for the detection of almost invisible elements: (a) Detected sockets and light switches with a probability value (b) Textured geometry with the detected sockets and light switches

2.2.2.3 Digital Preservation Service

The Digital Preservation Service is responsible for creating an information package that can either be submitted to the digital preservation system hosted at TIB or to an arbitrary (local) directory or other arbitrary storage system.

The information package contains the actual 3D models (e.g. IFC and E57 files), technical and descriptive metadata, semantically enriched data and information about geometric enrichments. The services described in the sections before are responsible for creating the content of the information package and provide the possibility to edit that content. The Digital Preservation Service then stores this package to an appropriate endpoint.

The service supports two output formats:

- Packages which are stored in the OAIS compliant digital preservation system "Ex Libris: Rosetta" hosted at TIB
- Packages transferable to an arbitrary OAIS compliant system via a BagIt package.

Currently the service supports the creation of information packages for the Rosetta systems. The development of the general container format "BagIT" has started and will be available in next versions of the service.

Software artefacts

Name	microservice-sipgenerator
Type	Web-API
SW-Component	SIPGenerator
Source	http://github.com/DURAARK/microservice-sipgenerator
Responsibility	WP2

Name	SIPGenerator
Type	Software component
Source	Closed source license
Responsibility	WP6

2.2.2.4 Search & Retrieve Service

The Search & Retrieve Service provides stakeholders with the possibility to search for archived data and to examine the retrieved data sets. Depending on the use case and/or the stakeholder, the search criteria may differ. E.g. a building owner may want to retrieve an archived model by specifying the address of the building. For urban planning, a user may want to select a geographical area on a map to retrieve all archived models within this area. This service provides the user with a customisable search interface where queries can be build and stored for reuse.

Software artefacts This Search & Retrieve service is using the "microservice-sda" and the "SDA" component as search endpoint, which are both used also in the Semantic Enrichment service. Their description is repeated here for the sake of completeness.

Name	microservice-sda
Type	Web-API
SW-Component	SDA
Source	http://github.com/DURAARK/microservice-sda
Responsibility	WP2

Name	Semantic Digital Archive (SDA)
Type	Software component
Source	http://github.com/DURAARK/sda
Responsibility	WP3

The screenshot in Figure 2 shows the graphical user interfaces for the Search & Retrieve service within the WorkbenchUI.

2.2.3 Python-bindings to Geometric Enrichment Services

The Python-bindings for accessing parts of the geometric enrichment services (described in Section 2.2.2.2) are a complement to the Service Platform regarding the integration of functionality into commercial (and also non-commercial) products in the field of architecture. We are providing a Python interface to the tools for easy integration into products like McNeel Rhinoceros[15], Grasshopper[16], AutoDesk Revit[17], Dynamo[18] or ArchiCAD[19].

Those products provide the possibility to integrate custom functionality via defined programming language interfaces (Python[20] in our case). We decided to not integrate DURAARK's geometric enrichment functionality via the Service Platform, which is possible in theory but does not perform well for that purpose as access to low-level geometric properties is needed. A web interface for this access would be too slow because of the involved network communication necessary when using the Service Platform. Therefore we additionally expose this functionality via a Python interface.

The Python Bindings have been developed for the Grasshopper3D environment, see screenshot in Figure 9. The DURAARK python bindings are supplemented with additional components, which allow Rhino Grasshopper to handle point-clouds sufficiently. These components provide the services necessary to prepare the data for the DURAARK services, for instance to import, crop and subsample of point-clouds before or analyse and validate the output of the DURAARK services. The delivery of the services as individ-

[15] https://www.rhino3d.com/
[16] http://www.grasshopper3d.com/
[17] http://www.autodesk.com/products/revit-family/overview
[18] http://dynamobim.com/
[19] http://www.graphisoft.com/archicad/
[20] Python Language: http://python.org

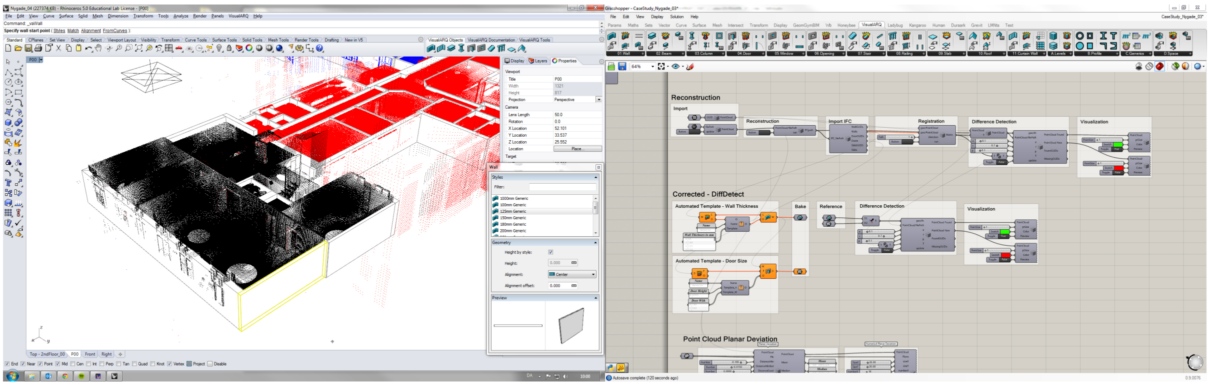


Figure 9: Integration of BIM and DURAARK functionality in Grasshopper/Rhino

ual components allows stakeholders to adapt the workflows to their business and project specific needs.

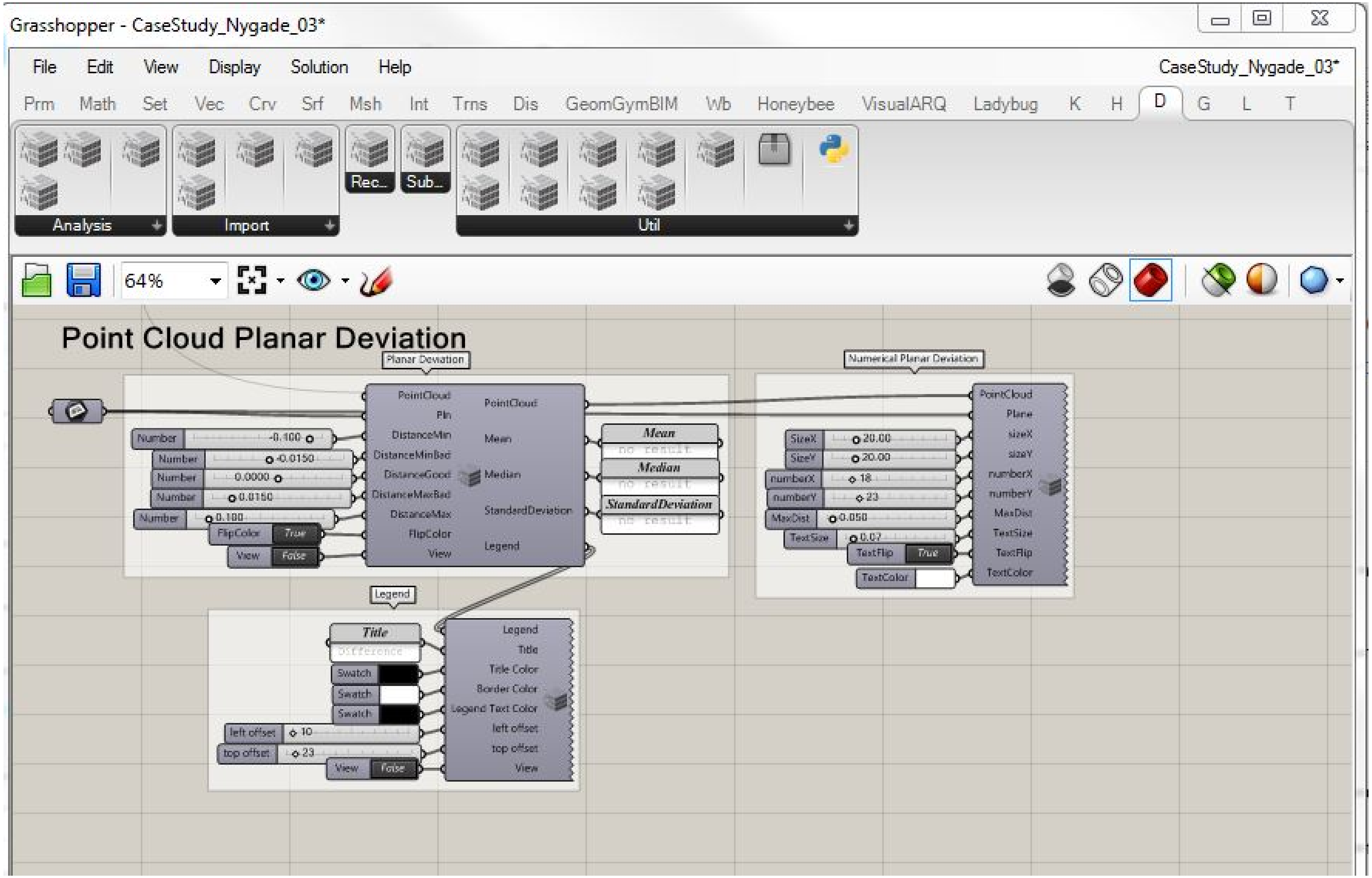


Figure 10: The DURAARK Python bindings are revealed as a collection in Grasshopper. The icons will be customised in the final release.

Software artefacts The Python Bindings for the Grasshopper Graphical user interface for Rhino will be distributed in year 3 of the DURAARK project via the Grasshopper3D website[21], which is the prime point of contact for programmers and user in this community. The distribution will include all components developed for the Grasshopper platform. Through this approach the user of the Grasshopper environment receives all components as a collection in the user interface. A detailed description of the Grasshopper components can be found in D7.2.

Name	Import
Type	Grasshopper Component via Python Bindings
Available at	http://www.grasshopper3d.com/
Responsibility	WP7

Name	Allocation
Type	Grasshopper Component via Python Bindings
Available at	http://www.grasshopper3d.com/
Responsibility	WP7

Name	Query and Select
Type	Grasshopper Component via Python Bindings
Available at	http://www.grasshopper3d.com/
Responsibility	WP7

Name	Analysis
Type	Grasshopper Component via Python Bindings
Available at	http://www.grasshopper3d.com/
Responsibility	WP7

[21] http://www.grasshopper3d.com/

Name	Output
Type	Grasshopper Component via Python Bindings
SW-Component	DifferenceDetection
Available at	http://www.grasshopper3d.com/
Responsibility	WP7

2.3 Datasets

2.3.1 Introduction to available datasets

The DURAARK project has collected and produced a considerable amount of 3D datasets in BIM and point-cloud formats. While the focus was set on the collection of files in open standards (E57 and IFC), several files also exist in proprietary formats such as vendor specific point-clouds (`Bentley Pointtools`, `Faro Scene`) or BIM-formats (`Autodesk Recap`, `Autodesk Revit`, `Bentley Microstation`, `ArchiCAD`). These files can be used to test the derivation of the Open formats, but neither their conversion nor their preservation are within the scope of the DURAARK projects. The collection of datasets emanates from all phases of building design, ranging from the early conceptual design stages to construction documentation and facility management.

While the datasets were useful for directing the course of the DURAARK project and provided the consortium with real world data, to assess research methods and software prototypes, many of the files cannot be shared publicly as they contain data sensitive to the owners and the security of buildings. Often they are the results of considerable amounts of work and in many cases owners of the files hesitate to share these for free. An exception are datasets from public authorities, and academic environments which can be shared publicly.

These dataset will form the basis of a public repository of datasets for test purposes similar to

- the **'Stanford Bunny'** and other geometric test datasets extensively used by the Computer Graphics research and development community for the validation and comparison of algorithms[22]

[22] http://graphics.stanford.edu/data/3Dscanrep/

- the **Common Building Information Model Files and Tools**[23] and the **Open IFC Model Repository**[24] for the BIM research and standardisation community to test and validate files.
- the **Lehigh University Benchmark (LUBM)** [25] datasets for the evaluation of RDF triple stores and SPARQL queries

This dataset has a value outside of the DURAARK project as well, providing examples as well as test and validation data.

Additionally, a range of structured RDF datasets – Linked Data – and related vocabularies are provided as part of WP3 and in detail listed and described in deliverable D1.7.

In the following sections, an overview of the four main categories of datasets is provided that are made available by the consortium.

2.3.2 Point-cloud Datasets

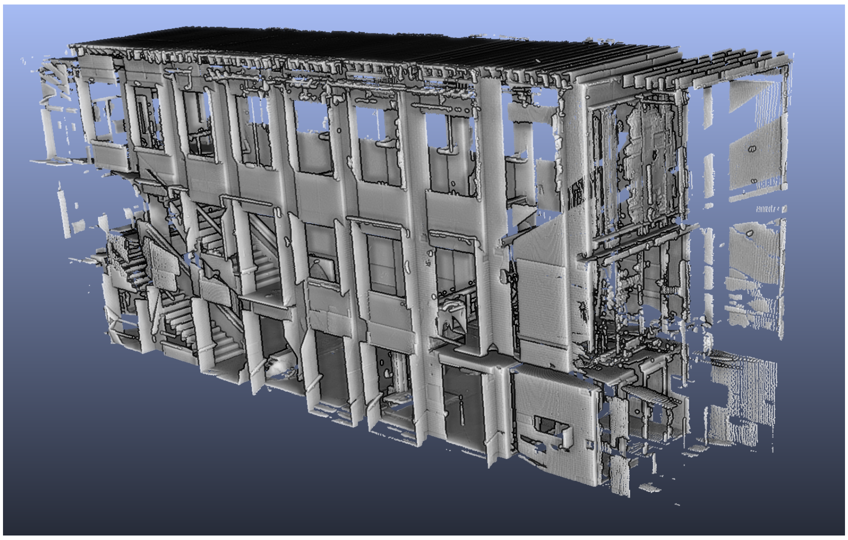


Figure 11: Screenshot of the small (500 MB, 22 million points)Acadia Office building scan made available as a E57 dataset.

[23] http://www.nibs.org/?page=bsa_commonbimfiles
[24] http://openifcmodel.cs.auckland.ac.nz/
[25] http://swat.cse.lehigh.edu/projects/lubm/

Confidentiality	Number of files	total size [GB]	average size [GB]
Confidential	61	373	6,01
Public	36	162	4,5

Table 2: Overview of public and confidential point-cloud E57 files in use and published.

A number of point-cloud datasets are being made available to the general public in the DURAARK project. From the scan datasets that have been collected from outside partners only a fraction can be made available to various security, privacy and licensing issues. However, many have been used for internal testing purposes and might be made available for other academic, standardisation and testing purposes upon request to the by the individual rights holders. Examples are provided in Table 3. Note that this category only covers E57 that have no explicit BIM representation side-by-side. The total number of available E57 is around 550 GB. A detailed overview of all datasets is provided in the Appendix A.3.

Dataset	Description	Author	size [GB]	Points $\times 10^6$
Acadia Office	Acadia Office Building	Acadia	0,5	22,58
CITA Room-24-7	Room 24 7 at CITA during the Workshop	CITA	1,0	41,78
DermoidI	Dermoid I installation description	CITA	1,3	63,50
DermoidIII	Dermoid III installation and calibration test	CITA	11,9	496,27
DesignHub	CITA Design Hub exhibition space	CITA	24,70	1021,95
Kronborg	Kronborg castle building	CITA	16,80	696,88

Table 3: Selected point-cloud datasets of varying size made available in the E57 format.

2.3.3 Building Information Model datasets

As highlighted in the introduction, the availability of common datasets for quality assurance and reference implementation in the fields of interoperability is of utmost importance. The DURAARK activities with regard to Building Information Modelling and

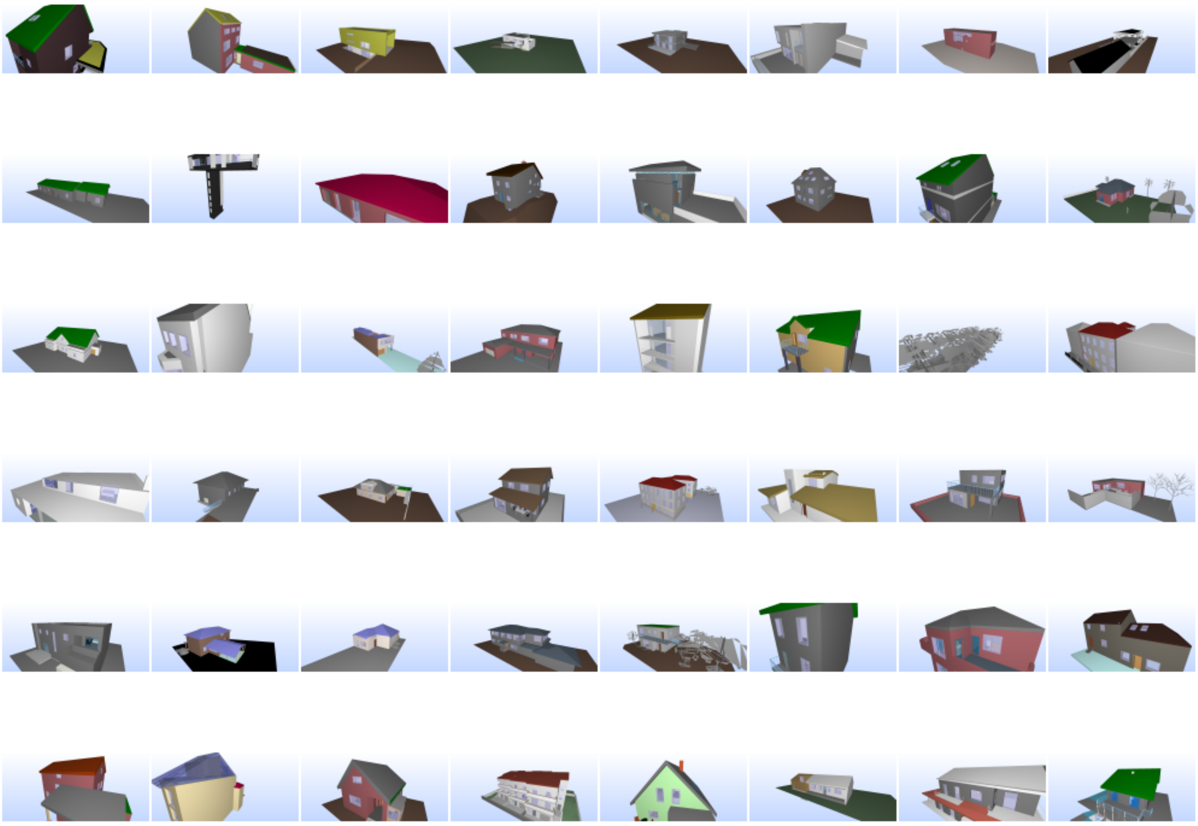


Figure 12: Partial overview of IFC example datasets provided by the DURAARK consortium.

particularly with regard to IFC models have been focused on reusing the existing established reference datasets and making new datasets available. Datasets listed here only have their explicitly modeled representations and mostly stem from planning and early design stages. A subset of the models is illustrated in Figure 12, an overview of the number is provided in Table 4 and the complete listings can be found in Appendix A.1.

Confidentiality	**Number of files**	**total size [MB]**	**average size [MB]**
Confidential	5	299	57,8
Public	210	2.286	10,9

Table 4: Overview of public and confidential IFC STEP Physical Files in use and published.

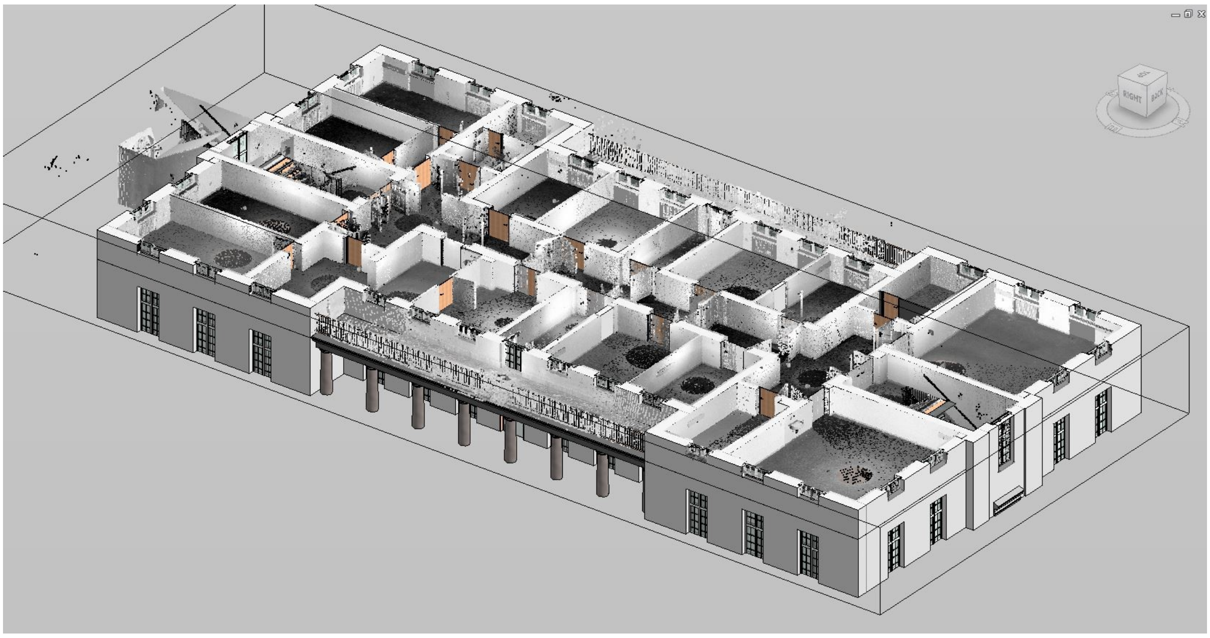


Figure 13: Chaircut through the hybrid "House 30" dataset that demonstrates the integration of the explicitly modeled building overlaid with the laser scan point-cloud datasets.

2.3.4 Hybrid point-cloud and BIM datasets

Hybrid models have both an explicit representation in the form of at least one IFC file and additional point-cloud datasets that have been acquired in the existing building. These datasets have been used in particular to test automated object recognition and registration of scanned data with modelled data. For the remaining activities these datasets will play an important role in the extension of the IFC model schema with point-cloud data sets (deliverable D3.5), where they will serve as test and validation datasets.

2.3.5 Semantic Building Metadata - Linked Datasets and Vocabularies

Semantic metadata for building models is produced in WP3, following Linked Data principles There are different categories of Linked Data and related RDF vocabularies that have been collected and produced in the context of the DURAARK project

- Vocabularies such as the `buildM` metadata schema[26] for the description of archived

[26] http://data.duraark.eu/vocab/buildm/

building models and scans as well as the actually described physical assets. The `buildM` vocabulary is the central schema for description of semantic metadata in the SDA. These cover technical as well as descriptive metadata.

- metadata of physical and digital assets in the form of `buildM` instances. The population has been gained by the extraction and enrichment of the BIM datasets described in 2.3.3. The population currently consists of ca. 5000 triples and is further described in the appendix.

- **Snapshots and profiles of datasets** from the Linked Data Cloud that have been gathered, compiled and generated by SDO components like the focused crawler (further details in the appendix).

- **Links and enrichments** of `buildM`-based metadata. This, on the one hand, includes **mappings between building-related vocabularies** such as the buildingSMART Data Dictionary (bSDD), the Getty Arts and Architecture Thesaurus (AAT), the FreeClass ontology that have been extensively described in the report D3.2. The interlinking curation and reinforcement is based on a number of automatically pre-alignments of some of vocabularies mentioned above. On the other hand, it includes links with external data, i.e. the semantic enrichments described in earlier deliverables (D3.2) and the ones generated through crawling, interlinking and clustering described in deliverables D3.4 and D3.6. Further details can be found in Appendix A.2.

3 Market study

This chapter consists of three parts; first a listing of the stakeholders, followed by a description of the different market sizes, and the last part contains Table 5, providing information about 12 potential markets that we think that DURAARK based tools could target.

Information for each market is summarised in Table 5, providing information about; market segment, the value proposition, the type of potential product/service, and a comment about the size of each market segment and the level of competition.

3.1 Stakeholders

This section describes stakeholders that have been identified and selected in the DURAARK project. These were defined and described in deliverable D2.1, while deliverable D7.1 investigated the stakeholders' processes with building-related information.

3.1.1 Building owners and facility managers

As described in deliverable D2.1, building owners and real estate managers are consumers of architectural data as part of the facility management process.

For this group, there exist different scenarios in which archived architectural data plays an important role. When intending to retrofit a building or when planning to erect a new one, developers and real estate managers usually address a municipal planning and building control office or a library with a collection mandate with a request for stored data.

Furthermore, when intending to sell a particular real estate, additional documentation of the buildings' history may be needed in order to assess its value. Risk assessment based on previous changes to the structure of a building as well as evaluating the potential to easily extend or refit a building are further interests specific to this group of stakeholders. In many cases, this information is not readily available in digital form. One way of acquiring the necessary information is by means of 3D scans.

We are specifically addressing **building owner** stakeholders through the SIP generation workflow in the DURAARK workbench. Being based on well-established metadata standards (METS and PREMIS), the preservation file format (IFC) – which today is seen as a good preservation format for BIM models – and further being based on best practices identified in the research presented in deliverable D6.1, the SIP generation workflow enables future-safe ingestion of BIM models.

Facility management stakeholders have expressed an interest in the possibility of searching and restructuring information extracted from BIM models or point-cloud scans. The enrichment components provided by DURAARK affords powerful alternatives to present handling, by broadening the possibilities for this type of search and restructuring.

3.1.2 Cultural heritage institutions

As described in deliverable D2.1, cultural heritage institutions are consumers of architectural data. Their task is to gather, preserve and grant access to the data specified in their collection profile.

Cultural heritage institutions include libraries, archives and museums at national, state or institutional levels. They are often responsible for information that has left the domain of industrial interest and needs to be preserved as part of the cultural heritage of a specific country or region. Many of these institutions have a clear archival mandate, often specifying the collection of material regardless of publication form.

Cultural heritage institutions have found themselves facing a growing amount of such digital content since the turn of the millennium. The UNESCO Charter on the Preservation of Digital Heritage (2003) defines the term "digital heritage" as "unique resources of human knowledge and expression", which must be made accessible while simultaneously respecting copyright and privacy rights. To meet this goal, it is necessary to develop and implement standards and strategies for digital preservation. As such it can be expected that cultural heritage institution will maintain long-term archiving solutions containing architectural data.

3.1.3 Research community

Considering the entire lifecycle of a building model and the different processing steps involved, the key problem areas and scientific disciplines of relevance to DURAARK are:

- Digital Preservation and Digital Libraries
- Building Information Modelling
- Data and Knowledge Integration, in particular, Semantic Web and Linked Data

In addition, secondary domains include the areas of *Computer Graphics*, *Human Computer Interaction* or *Machine Learning*.

3.1.4 Architects and engineers

As described in deliverable D2.1, architects and engineers are both producers and consumers of archived architectural data; while they consume architectural data as a base for their planning activities, they also want to archive the data they create, providing long-term access.

As data consumers, architects and engineers query a DURAARK archive in order to find information about a building's history and the changes it has undergone. Their focus depends on the job that is underlying their search and might vary from an investigation into the original construction of a building, from which they can derive the intention of the original author, to the investigation of the urban context of a potential building site or the assessment of the current state in spatial and construction related areas of the building represented through a relatively up to date 3D scan.

As the planning of buildings is done on an abstraction of reality, architects will typically ask for an abstracted model that represents the state of a building. Such models are typically-polygon based and stored in IFC format. A mechanism to derive these models from 3D point-clouds, as well as an indication on the deviations from the real geometry will allow architects to start their planning directly from the archived data. The enrichment of this data through existing data from previous planning states is expected to have a positive impact on the planning precision, not at least for the improvements of the energy use of building structures.

Architects and engineers can be commissioned to monitor and evaluate the state of a building in order to detect damages and the cause and speed of their progression. Archived descriptions of single buildings and aggregations are of high value here and the tracking of changes, caused for instance through underground building works or degradation of elements, is directing the engineers' recommendations.

As engineers and architects are the main producers of architectural data it can be expected that they will install or use long-term archiving systems for their own companies' architectural data. The motivation for this might be to document and secure the data that describes a project delivered for the period of liability, but as well the ingest into the companies' archives that in bigger organisations usually is more directed towards public relations, acquisition and internal knowledge management.

Engineering companies are usually bigger than those of architects, making the installation of long-term archiving systems even more probable.

3.1.5 Land surveyors

Land surveyors typically participate in the beginning of any planning process, whether related to a new building or a renovation. Their discipline is in its core related to the abstraction of the world's complexity to understandable information for users that want to process this information

Construction companies are handling all parts of the realisation of buildings including the related detailed steps in the planning process. Their responsibility is the accurate translation of the abstracted building intent into physical form.

Land surveyors typically participate in the beginning of any planning process, whether related to new buildings or renovations. Their responsibility is the accurate translation of the abstracted building intent into physical form. Based on this they would have great use of access to information about existing buildings.

3.1.6 Construction companies

As described in deliverable D7.1, construction companies are handling all parts of the realisation of buildings including the related detailed steps in the planning process. Their responsibility is the accurate translation of the abstracted building intent into physical form.

As described in deliverable D2.1, architects and engineers as well as construction companies can be both producers and consumers of archived architectural data. They predominantly produce highly detailed and annotated architectural data in the later stages of a building project. However, they also consume previous records of the building they are working on for evaluation purposes.

During the often years-lasting building projects, construction companies produce vast amounts of architectural data, which needs to be stored during the process of the project and beyond. This especially holds true as the construction companies, as the last part in the design to production chain, are today starting to utilise the full potential of building information models with architectural objects with rich metadata. Here the level of detailing is increasing and besides the modelling of very small objects in BIM models (from 5mm onwards), especially the integration of external catalogues is of importance. Such catalogues hold information about products from building vendors. The increasing specifications in BIM models are becoming part of the business model of construction companies.

As consumers, the access to a long-term archive will help construction companies with questions regarding measurements in existing parts of a building. The comparison of recent scans of an ongoing building project with those that are stored in the long-term archive will allow them to track the building process, e.g. see whether tolerances are met and building elements are installed in time.

As many building companies today are trans-national they will have a natural interest in building up their own long-term archiving systems, especially as they are highly aware of the legal aspects of their work.

3.1.7 Public administrators

It is likely that it will become normal in some European countries that building authorisation applications can be sent to the legislature utilising BIM files. In order to handle this information both long-term digital preservation and semantic technology are relevant.

3.1.8 Software companies

One important type of player in the market is software companies providing tools for the construction industry. This includes huge companies like AutoDesk and small companies like the DURAARK project partner Catenda. They might want to integrate software artefacts created in DURAARK into their software tools. This could be in the form a middle-ware components or end-user oriented software products.

In many cases the end user companies, like the architectural and engineering companies, are not primarily interested in middle-ware components, since the majority of them will not develop software by themselves (but some do). In those cases software companies might be the vehicle for getting DURAARK content in the hands of the end users.

3.2 Size of markets

3.2.1 Size of the laser scanning market

A study[27] that was published in July 2013 states as a forecast that the market value related to 3D scanning will grow from USD 2 billion in 2013 to USD 4 billion in 2018. This number includes devices and services in all markets – everything from cultural heritage to defence. This is a commercial study and expensive to get access to, so only the public part of the report was available to us. This means that the breakdown between different regions and sectors are unknown.

Another study[28] from July 2014 estimates that the world market value will be USD 4.9 billion in 2020. This number seems very compatible with the the 2018 number from the previously mentioned study. It also states that the largest market is expected to be Europe, with a revenue of USD 1.65 billion in the same year.

3.2.2 Size of the construction industry

According to Eurostat[29] there were about 823,000 construction enterprises in EU ("EU 28") categorised as being in the sector "Construction of Buildings". The largest numbers

[27] http://www.marketsandmarkets.com/Market-Reports/3d-scanning-market-1110.html
[28] http://www.alliedmarketresearch.com/3D-scanning-market
[29] http://www.eurostat.ec.europa.eu

were in Spain (163,000) and in Italy (142,000). However, many of these are very small. The total EU number of enterprises with less than 10 employees ("micro enterprises") is 770,000. So most construction companies are in the smallest category. This is important background when considering the market potential that these companies represent. The reason is that smaller companies might be less willing to invest time and money in software. On the other hand the larger companies are more likely to develop a tailored solution in-house instead of purchasing a solution.

Figure 14 shows the size of the construction industry compared to the other sectors in the EU economy. It confirms that it is one of the largest sectors in the EU economy. It is common knowledge that it is one of the largest sectors in all industrialised economies. Figure 14 also tells us that the sectors relative size is larger when measured in labour than when measured in the value added. This might be an indication that the sector have a tradition of low investments.

Figure 15 shows the construction index from 2003 to 2013, with the index normalisation value of 100 determined by the value in year 2010. It clearly shows the large shift related to the financial crisis. This is important since the willingness to invest might be lower until there are periods of sustained growth in the sector.

3.2.3 Size of the BIM market

A market study[30] from October 2014 predicts the BIM market to grow with 17.4 percent annually in the time period 2013-2018. This report states that one of the major challenges in this market is the lack of expertise and the high cost of training.

Another market study[31] from 2012 predicts the BIM market to reach USD 6.5 billion worldwide by 2020.

A study[32,33] by consultancy Competitive Advantage from 2013 states that "at present" (2012) BIM is used for 3.9 percent of all construction projects representing EUR 3.8

[30] http://www.alliedmarketresearch.com/3D-scanning-market

[31] www.navigantresearch.com/newsroom/building-information-modeling-market-to-reach-6-5-billion-worldwide-by-2020

[32] http://www.themanufacturer.com/articles/firms-need-bim-to-access-a-27bn-market-by-2016

[33] http://cadvantage-knowledge.co.uk/#!/~/product/category=8803218&id=28854052

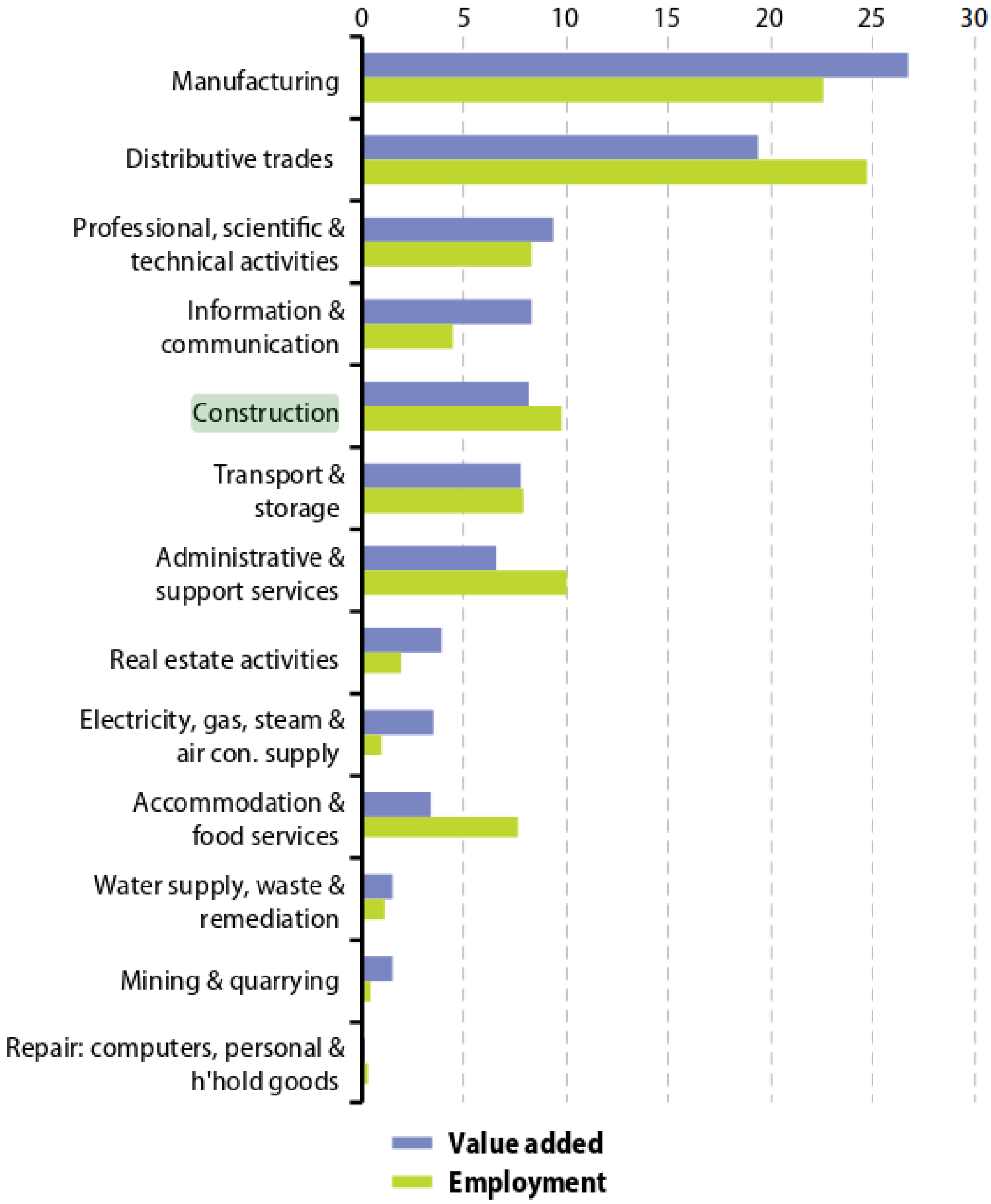


Figure 14: Size of the economic sectors in 2011 measures by value added and employment. Source: Eurostat.

billion. But it predicts BIM penetration to have increased to 50.8 percent of projects – worth EUR 55.1 billion (measured by the total value of construction projects where BIM used).

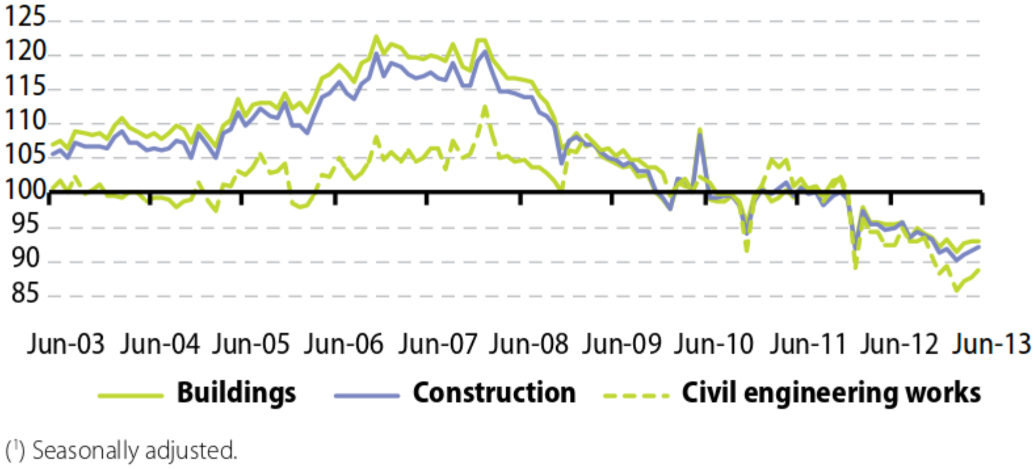


Figure 15: Index of production, construction, EU-28, 2003-2013.
Source: Eurostat (online data code: sts_copr_m).

McGraw Hill Construction (now Dodge Data & Analytics) published a report[34] in 2013. The report states that from 2007 to 2012, BIM adoption in North America jumped from 28 percent to 71 percent, while the region's contractors recently passed architects on the adoption curve (74 percent, as compared to 70 percent for the latter group).

The latest news at the time of writing (February 2015) is the statements[35] from the governments in France and Germany. They are pushing for larger adaption of BIM in the building, construction and facility management industries.

All in all there are strong indications that the BIM market is large and that it is rapidly increasing. This will probably be accelerated by governmental initiatives like the one mentioned in the previous paragraph.

3.2.4 Size of the long term digital archiving market

IDC published a study[36] showing that the market for Digital Archiving in 2010 was USD 1.5 billion worldwide, and that it grew 13.6 percent from 2009.

[34] http://www.aconex.com/blogs/2014/01/global-state-of-bim-construction-market-data.html

[35] http://bim.construction-manager.co.uk/news/france-and-germany-move-forward-bim-adoption

[36] http://www.slideshare.net/arms8586/idc-archiving

There are some other studies too, but they appear to be older. What many of them have in common is a focus on storage media (tape, hard drive etc) and cost per byte of storage.

A newer study[37] (2014) from Gartner looks at "Structured Data Archiving" and estimates the market value to be USD 270 million last year (2013), but that it is growing 10 percent per year. The study describes the status for each of the large players.

This is very relevant for DURAARK, and even though our project has a focus on a niche (but potentially a large niche) the status of the big players as described in the Gartner report is very important to be aware of.

3.3 An overview of market segments

In Table 5 below, an overview of market segments and important aspects for evaluating potential profitability is presented. The estimates have large associated uncertainties, and further studies would definitely be necessary before taking steps towards starting a business venture.

In each of the rows listed, the column "value proposition" suggests which of the DURAARK artefacts that are thought to be most relevant for the specified market.

We have not attempted to estimate the monetary value of each of these markets more precisely, falling back on a broad estimate of their relative size by terms such as "small" and "large". The reason is that a more precise estimate would be very difficult to do, especially before the value proposition is further developed – and a large part of this will be the result of further work in the DURAARK project in it's third and last year.

[37] http://www.gartner.com/technology/reprints.do?id=1-1VDEQU0&ct=140611&st=sb&mkt_tok=3RkMMJWWfF9wsRonua3Bc0%252FhmjTEU5z16O0pUa6zgIkz2EFye%252BLIHETpodcMS8NjNa%252BTFAwTG5toziV8R7HNJc160s8QXBjm

Id	Segment	Value proposition	Potential product/service	Market size	Competition	More info
M1	Building owners and real estate manager	Long-term digital storage	Archiving product with related services	Considerable	Low (fragmented)	Section 3.1.1
M2	Building owners and real estate managers	E57 to IFC production with intact references back to E57	Middleware	Small	Several	3.1.1
M3	Cultural heritage institutions	System for their building-related content	Product or service	Small in each country, but each country has at least one of these	Medium size	Section 3.1.2
M4	Facility managers	System for their building-related content	Product or service	Big	Large (but very diverse)	Section 3.1.1
M5	Research community (as customers)	System for their building-related content	Software-as-a-Service (SaaS) or middleware components?	Small	Medium (but very diverse)	Section 3.1.3
M6	Architects and engineers	Archive the data they create, providing long-term access	SaaS	Large	Fragmented competition with very different approaches	Section 3.1.4
M7	Architects and engineers	Extract IFC from point-cloud	SaaS or middleware components	Large	Relatively few, but many will probably try to enter this market	Section 3.1.4
M8	Land surveyors	System for their building-related content	SaaS or middleware components	Large	Unknown	Section 3.1.5
M9	Construction companies[38]	Archive the data they create, providing long-term access	SaaS	Large	Large (but very diverse	Section 3.1.6
M10	Construction companies[39]	Document and deliver "as-built" with IFC and E57	Middleware components	Potentially Large	New market with large potential very diverse	Section 3.1.6
M11	Public administrator	LDP and semantic search	Middleware components	Large	Existing administrative software	Section 3.1.7
M12	Software companies	Require software to proved new features to end-users (related to LDP, IFC and E57)	Middleware components	Potentially Large	Mostly in-house teams at the software companies	Section 3.1.8

Table 5: Definition and description of market segments

[38] Require tools to access and query large amounts of IFC and E57 data automatically
[39] Require tools to validate and enrich data (IFC and E57)

4 General exploitation and sustainability strategy

4.1 Exploitation & sustainability goals

The main goals of the DURAARK project is to develop methods and tools for long-term preservation of architectural knowledge in order to accommodate future access to this information, but in a manner that accommodates future re-use of the preserved 3D models, that enriches building information models with "as built" information from point-cloud scans, and that allows semantically enriching building information models with additional data sets.

Knowledge about buildings and built structures is of interest to a wide variety of stakeholders, ranging from architects and urban planners to building operators or the general public. Such knowledge includes 3D models and point clouds as they are generated throughout the planning, building, construction or refurbishing phases. Related information about the legal, historical, infrastructural or environmental context of built structures is considered useful in many use cases in practice. While such information as well as the actual structure evolve continuously, preservation of architectural knowledge is of crucial importance.

The tasks that we aim at carrying out after the project's end can be grouped in these main areas:

- providing long-term access to DURAARK results
- stimulate take-up and reuse of DURAARK results (see section 2)
- enabling third parties to benefit from DURAARK results
- growing the user base and the community built around the DURAARK project
- providing DURAARK results as a foundation for business models (from third parties or consortia)

In preparation for the project's third year, discussions about sustainability and maintenance of the project outcomes were initiated. Part of our strategy is to initially release our results as openly[40] as possible to the outside world – we are trying hard to create a community around the outcomes that we are later going to exploit in preparation of a

[40] "Open" in the sense of licensing

robust business model/business idea. So, our strategy for the immediate future after the project duration is to make outcome available as much as possible. Where applicable, we have therefore chosen open licensing models where we aim at releasing as much as possible as openly as possibly.

Based on this reasoning, our very first steps are to establish a thriving community around the DURAARK results in preparation for business building steps. Building a business includes defining and establishing a business model as well as identifying a market. Our efforts into disseminating information, releasing datasets and software artefacts, and maintaining these over time – steps the consortium is already taking – is not interfering with building a business around the outcomes.

We observe that many successful start-ups and internet businesses emerge from scenarios where they initially gather a thriving community before trying to commercialise on this audience. The consortium therefore decided to invest in attracting and reaching a group of people, building a community and a larger audience before establishing a legal entity. Only after having assembled a critical mass of users and interest do we find it feasible to invest in creating a legal entity to exploit the fact.

4.2 Exploitation actions overview

Elaborating on the tasks that we plan for after the duration of the DURAARK project:

Providing long-term access to DURAARK results after the life-span of the project, is a necessary precondition for all (other) activities. We are currently looking into using a commercial preservation service (APA and APARSEN[41]) to ensure future access to reports and datasets. Software artefacts emanating from the project have already been added to GitHub, at `http://github.com/duraark`, where they are readily available together with documentation that covers building procedures.

Stimulating take-up – and later reuse – of DURAARK results is work in progress. The project is actively promoting itself through the organisation of workshops with the earlier mentioned user categories and with third-party companies that provide services around building information models.

[41] http://www.alliancepermanentaccess.org

Enabling third parties to benefit from DURAARK results is addressed by opening up datasets and software artefacts and publishing them through channels such as GitHub.

Growing the user base and the community built around the DURAARK project is addressed by actively disseminating information during the life time of the project, by organising workshops, and by participating in existing communities such as the W3C Community Groups and the BuildingSmart community.

Providing DURAARK results as a foundation for business models is addressed in a similar way as above, e.g. by building and expanding a community.

4.3 Management of foreground

The intellectual property (IP) rights to *foreground* produced within the DURAARK consortium during the project's lifetime, are discussed in deliverable D1.7 and regulated by the Consortium Agreement (CA).

The DURAARK consortium has these processes in place, regarding management of foreground:

- the consortium is actively monitoring existing IP (as laid out in deliverable D1.7)
- the consortium is actively on look-out for new IP arising from the work in DURAARK (which is added to deliverable D1.7)
- the consortium is actively looking for ways to exploit and disseminate project results (as laid out in deliverable D8.6 and this deliverable)
- the consortium is monitoring IP-related issues that were not foreseeable at the start of the project, specifically looking into issues related to IP jointly generated by partners – for instance by drafting joint ownership agreements for specific cases, if this need should arise

In preparation for the conclusion of the DURAARK project, we are looking into an agreement around foreground that have (or may have) commercial potential. In particular, we are planning for:

- how to exploit the ownership (mainly discussed in this deliverable)

- how to share IP costs (if need arise)
- how to share revenues
- the appropriate license and assignment rights (deliverable D1.7) – possibly also covering confidentiality of specific information, such as source code

Particularly, issues related to transfer of ownership has to be taken into account before the consortium dissolves. As we have an aggressive licensing approach, choosing as open licenses as we possibly can, we may not have to consider assigning ownership rights to those communities (working groups) that we are contributing to[42].

The crucial point to take into consideration here is the direct industrial applicability of the outcome of the DURAARK project (discussed in section 2), which heavily influences the question of future shared revenues. As a general rule, the protection of rights we discuss should reflect the use we wish to make of the foreground. In preparation for scenarios such as if a partner finds an opportunity to make money out of the results of the project, we need to have an agreement in place that regulates how the IP rights apply and how revenue should be shared.

Examples of commercial exploitation:

- offering of consultancy around DURAARK results,
- offering of services based on the DURAARK framework, e.g. providing preservation services specifically targeted at BIM data,
- offering of services based on individual components or APIs produced by DURAARK.

We do still point out that our current focus is on growing a community (user base) around the outcome of the project, rather than focusing on an immediate commercial exploitation. The rationale for this approach is further discussed in section 5.1.

The consortium understands that if foreground capable of industrial or commercial application has not been protected, the Commission may protect it on behalf of the EU[43]. We are currently covering all foreground with licenses of various degrees of openness (see deliverable D1.7).

[42] In case ownership is transferred, the new owner must accept the obligations we are putting forth in this deliverable regarding dissemination and use

[43] Article 44.2 of the Rules for the Participation in FP7 projects and Article II.28.3 of FP7 Grant Agreement – Annex II

4.4 Planning ahead

In light of our primary approach on growing a thriving community around the outcomes of the project, the DURAARK consortium proposes to start a European COST Action[44] in order to accommodate a network of researchers as well as practitioners in the area where DURAARK is currently active – providing durable architectural knowledge by addressing digital preservation of, and the use of, geometrically and semantically enriched building information models. This would fit well within the Information and Communication Technologies domain, addressing the research areas of "Information science and technologies", as well as "Societal aspects of ICT".

Utilising existing networks already established through DURAARK, or through other connections, will serve as a foundation for setting up an initial COST Action consortium. The DURAARK consortium as such would include six COST Countries, thereby making it eligible for application, but connections such as buildingSMART[45] and DIMMER[46] within the BIM sector, and e.g. DLM Forum[47] within the digital preservation sector, would be valuable in reaching out wider, especially geographically, while still remaining in the scope of the proposed COST Action. Although a COST Action does not fund research as such, it provides means for networking in different ways, e.g. workshops, training schools, conferences, and short-term scientific missions (i.e. visiting researchers), which is valuable in the sharing and application of knowledge and ideas. This would accommodate an open and inclusive environment, where new members can join at any time.

Collaboration around the COST Action application will be lead by LTU and the aim is to submit the application before the autumn collection date (not specified yet, but it is envisaged for September 2015).

[44] http://www.cost.eu/
[45] http://www.buildingsmart.org/
[46] http://dimmer.polito.it/
[47] http://www.dlmforum.eu/

5 Exploitation and sustainability plan

5.1 Organisational sustainability and community-building

The consortium is currently discussing the forms of an organisational platform for utilisation of project results. There are a number of concerns that have to be taken into consideration, ranging from how best to attract new participants, to the economic feasibility of building a successful business.

Regarding attracting new participants, a legal entity might be seen as too much structure and overhead for getting new participants to join. Based on our interactions with the relevant communities, it is more appealing for individuals to join working groups or special interest groups (SIGs).

Some efforts have been invested in considerations around establishing a new legal entity – in part by taking care of activities such as maintaining datasets and software artefacts. We have discussed possible legal forms from which it may be reasonable to choose. One important consideration has come from this effort; many legal forms do not match the requirements of a community effort such as DURAARK. At the same time, DURAARK is already affiliated with existing communities that in many ways provide the benefits of a legal entity while at the same time avoiding the additional cost and hassle with setting up a new legal entity.

Additionally, the consortium is leaning towards building on existing communities and organisations that in many ways have the same goals as DURAARK, rather than establishing a new entity from scratch. We currently believe that it is more efficient and realistic to actively combine efforts with an existing community, rather than pursuing a competing path with community building around a new legal entity.

Considering our primary approach to grow a thriving community, setting up a COST action (see section 4.4) would be a useful instrument to build endurance into our effort.

5.2 Standardisation

The ongoing and future standardisation efforts within the DURAARK project are executed in three strategic areas that are tightly coupled to ongoing community-building efforts.

The three strategic areas are:

1. Standardisation for the **building and construction industry** in the buildingSMART organisation

2. Standardisation of semantically enriched Building Information Models in the **Linked Data and Semantic Web community** represented in the W3C organisation

3. Standardisation of technical and descriptive metadata about building for the **Archival Community**

In the following sections these standardisation activities will be described in detail.

5.2.1 Standardisation in the buildingSMART organisation

The buildingSMART organisation is an international standardisation body focused on specification of interoperability agreements in the fields of Architecture, Engineering, Construction and Facility Management (AEC/FM). Established as a spin-off of the ISO 10303 family of standards referred to as STEP (STandard for the Exchange of Product models),the organisation currently hosts the most relevant open interoperability standards in the building and construction industry:

- The Industry Foundation Classes (**IFC**) data model which has been extensively described in earlier deliverables across all work packages. IFC instances are the main information carrier for long term preservation of Building Information Models in the context of the DURAARK project[48].

- The International Framework for Dictionaries (IFD, ISO 12006) as concept and vocabulary structure originally intended for the homogeneous organisation of the various classification standards. A reference implementation and population of concepts officially hosted and supported by the organisation is the buildingSMART

[48] besides E57

Data Dictionary (**bsDD**). The development of the framework as well as the production servers are currently hosted by DURAARK consortium partner Catenda.

In the context of the DURAARK project, the bSDD is extensively used for the semantic enrichment of ingested models and their metadata. Detailed descriptions of the framework itself as well as the enrichment can be found earlier Deliverables of WP3.

- The Information Delivery Manual (**IDM**) and Model View Definition (**MVD**) standards for the structured, process-based exchange of data models and further requirements formulations. For the formal descriptions of information required during the exchange of information for the purpose of archiving, experiments have been conducted in the context of the DURAARK project to develop a dedicated MVD for archival purposes similar to e.g. the PDF/A profile.

  While theoretically desirable and technically feasible the effort has been considered unrealistic in current business practice. The preliminary results however will be submitted to the buildingSMART organisation for future standardisation efforts.

- The Building Collaboration Format (**BCF**) for the formalised exchange and management of issues (e.g. defects during construction, model clashes etc.). In future DLP scenarios this could be used in the context of reporting checking results based on and 'IFC/A MVD' during the ingest of IFC files.

The formal procedure to propose new standards to buildingSMART includes significant promotion work in the community, gathering of support and presentations during the international technical summits including the respective sub-groups and committees of the organisation such as ISG (Implementer Support Group), IUG (International User Group), MSG (Model Support Group), TAG (Technical Advisory Group) and – most importantly for technical specifications – ITM (International Technical Management).

Throughout the DURAARK project such promotion of the foreground has been done contentiously during several meetings. To increase the success of the standardisation efforts, the highest ranking technical buildingSMART member (Thomas Liebich, AEC3[49])

[49] AEC3 is a BIM consulting firm for the building sector, based in Munich, Germany

has been successfully invited to the DURAARK board. During DURAARK meetings and workshops the DURAARK progress and (intermediary) results have been presented on several occasions and feedback has been incorporated into succeeding R&D efforts. The contributions of the DURAARK project to the standardisation efforts of the build-

ingSMART organisation are focused on a number of areas with varying degrees of impact which are described in the following overview:

Exposing and maintaining bSDD as Linked Data and mapping it to other vocabularies. For the preservation of semantically rich models, the use of external references from within IFC instances is expected to rise in the future. To allow a complete preservation of the entire model including its external references, the choice to transform the bSDD into RDF has been made in the context of the DURAARK project.

The resulting RDF version of the bSDD has been presented to the bSDD working group during the buildingSMART international meeting in Stockholm in 2014. Since then it has been used by the community for a number of mapping processes, duplication detections, performance evaluations and a number of experiments.

The complete endorsement of the RDF version of the bSDD is currently held up by conflicts with potential business models of the buildingSMART organisation. Currently, the visions favoured by the leaders of this community is to re-finance the development through use fees of member organisation chapters and individual businesses and stakeholder that would like to use the vocabularies.

If published and exposed under permissive licenses like in the context of the Linked Open Data movement – as is strongly suggested by the DURAARK project and other members of the community – these business models would have to be reconsidered and re-arranged. Note that licensing and pay-per-use models are also technically suitable for Linked Data. The DURAARK consortium will continue to promote these approaches throughout and beyond the lifecycle of the project. To this end, members of the consortium regularly participate in the bi-weekly conference calls of the bSDD working group.

Integration of (compressed) point-cloud structures in future versions of the IFC core schema. Although only starting in year 3 of the DURAARK project, the efforts envisaged in the project description and DOW to develop a point-cloud extension to the existing core schema of the IFC model has been promoted in the buildingSMART community even before the official start of the DURAARK project during an International Technical Meeting in Tokyo in November 2012.

It was embraced unanimously by the community since it has been a long-standing item on the agenda and strategic visions for the buildingSMART organisation. Early conceptual approaches have been discussed already and preliminary experiments have been carried out. However the peak of these efforts comes in year 3.

Standardisation and promotion of the use of RDF serialisations of (distributed) IFC models as Linked Data. Although it may not reach a broad scale use within the next few years, capturing and storing building information models as Linked Data is on the strategic agenda of the buildingSMART organisation. The topic has also frequently been identified as a future item in many research roadmaps (including those of the European Union) and in scientific literature. newline
A number of attempts to add RDF/OWL serialisation on a generic level spanning all industry domains within the STEP initiative, similar to the part 28 of the ISO 10303 (which specifies the serialisation of STEP models in XML data formats) have been undertaken over the years[50] but never managed to gain significant traction in the respective standardisation committees. newline

Partially, this can be attributed to the very early effort that was made before RDF and Linked Data had gained wide-spread acceptance across many IT application and interoperability domains. Early attempts to promote such approaches have been made by DURAARK consortium members in the context of the buildingSMART standardisation community as well as in the EU FP6 project "Inteligrid".

Results, demonstrators and use cases have been developed further over the years including the examples developed in the DURAARK context and documented in WP3 deliverables D3.1 and 3.3. DURAARK results are contributing to the efforts in a buildingSMART working group that will be proposing a definite ifcOWL profile

[50]examples include the http://exff.org initiative

to the International Technical Management during the summit held in March 2015 in London, UK. This will be a significant step towards both facilitating concurrent, collaborative engineering and sustainable digital long term preservation using open formats.

Model View Definition for archival purposes. Developments towards creating Model View Definitions / information profiles for archival purposes will be formally submitted as an MVD specification working group to the respective buildingSMART platform. In related research and development, partners of the DURAARK consortium have created and tested the first automated checker for formal specifications of MVD requirements in the mvdXML format.

The prototypical software has been published as Open Source and has been circulated in the buildingSMART community. As mentioned in the introduction of this section however the wide-spread support of MVDs apart from the most established ones (e.g. Coordination View (REF)) cannot realistically be expected in the immediate future.

5.2.2 WC3 community group for Linked Data in Architecture and Construction

Supported by the workshops organised and co-organised in the context of the DURAARK project and a broad discussions in a community communicating over a number of different channels such as mailing lists and social media platforms, a W3C community group for "Linked Data in Architecture and Construction" has been formed.

The group aims at standardisation efforts for the support of Linked Data technologies in the building and construction sector. As the first primary goal, the submission of a recommendation of a common notation and profile for an OWL/RDF version of the Industry Foundation Classes is under active development. These efforts are carried out in parallel to comparable efforts within the buildingSMART community. Next to members of the DURAARK consortium active members are involved in other European research and development projects such as DRUM and ready4smartCities.

5.2.3 PREMIS

The editorial committee of the de-facto preservation metadata standard PREMIS hosts PREMIS Implementation fairs, which usually take place in conjunction with the iPRES conferences. The DURAARK project presented the work of the pre-ingest project at the last PREMIS Implementation Fair held in conjunction with the 11th International Conference on Digital Preservation – iPRES 2014.

As part of the presentation, questions associated with the nature of archival processes around architectural 3D data and the integration of the pre-ingest workbench into existing digital preservation systems were put forward and discussed. The forthcoming PREMIS v3 will be addressing some of these issues and the DURAARK work is regarded a well-suited use case for the changes included in the new version[51].

5.2.4 PRONOM profiles for E57 and IFC-SPF

For the standardised identification of archival data, signature patterns for specific file formats are used and registered across implementation borders of individual LTP tools in the PRONOM registry. In this central registry both PRONOM Unique IDs (PUIDs) per file format and the respective signature patterns (similar to the Unix `file` utility) are maintained.

The absence of such profiles for IFC and E57 files has been identified as part of the gap analysis carried out earlier in work packages 2 and 6. A successful standardisation effort of the DURAARK project is the creation, submission and acceptance of such profiles – e57 (`fmt/643` and IFC-SPF (`fmt/659`) – to the PRONOM registry, which now can be reused in an interoperable and hence sustainable tools in future archival tools for the long term preservation of building information models and point-cloud scans. Further details can be found in deliverable D6.2.

[51] Source – McKinney, P., Zierau, E. and Guenther, R. PREMIS Implementation Fair Workshop. In: Proceedings of the 11th International Conference on Digital Preservation. 2014. pp 306-308.

5.3 Release of data, software, services – Licensing and IPR management

One of the central aims of DURAARK is to make project results as available and accessible as possible for third parties (see Section "Exploitation Goals"). This can only be achieved by (a) providing the legal and licensing conditions under which take-up and reuse is possible, (b) following open standards and technologies wherever possible and (c) releasing results on widely established platforms. With respect to (b), i.e. the use of open standards, DURAARK embraces widely accepted technologies such as REST for open APIs and services, or widely established Linked Data principles for data and metadata.

While DURAARK aims to reach large communities and audiences, results are not only widely shared on the project website, but also disseminated through established community platforms such as `GitHub`[52] or the `DataHub`[53]. In order to not just release DURAARK artefacts, but to also provide the legal and licensing environment in which wide adoption and reuse is facilitated, DURAARK is continuously identifying suitable licensing options for software artefacts and documentation produced during the lifetime of the DURAARK project, as laid forth in deliverable D1.7.

We consider three main types of generated foreground artefacts subject to IP protection: *reports*, *software*, and *datasets*, where the default intended licensing schemes are presented as follows. While this represents a general strategy, on a case-by-case basis, deviations will be necessary, for instance, to accommodate the licensing requirements of third-party software libraries.

- **Reports:** For reports of all kinds, the preferred license scheme is adopted from the Creative Commons[54] license scheme. In particular, we suggest the **CC Attribution** or **CC BY** license, which grants permissions to share, copy, distribute, and transmit the work, and also allows to make commercial use of the work, provided that the work is attributed in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

[52] `http://github.org` – for access to source code
[53] `http://datahub.io` – for access to datasets
[54] Creative Commons – `creativecommons.org`

By default, our strategy is to use the license **Attribution + No Derivatives** or **CC BY-ND**. **CC BY-ND** prescribes that the work may not be altered, transformed, or be used to build upon.

The full text of the license is available at Creative Commons[55].

- **Software:** While DURAARK intends to make all software publicly available and enable reuse by third parties, the suggested schemes adopt Open Source Licensing principles, for instance, as approved by the *Open Source Initiative – OSI*[56]. In particular and by default, we intend to use the GNU Lesser General Public License or LGPL[57]. The LGPL allows developers (e.g., in academia and companies) to use and integrate LGPL software into their own (even proprietary) software without being required to release the source code of their own software-parts. This represents a compromise between the strong *copyleft* of the GNU General Public License or GPL and permissive licenses such as the BSD licenses and the MIT License[58]. The choice for contextual LGPL license in DURAARK has been made as a middle ground between two main interests:
  - Reciprocal licenses like GPL v3 would have prevented downstream exploitation and commercialisation by demanding full disclosure of the derivatives. In particular for the SMEs directly involved as a consortium partner or as associated partner, such licensed would be problematic
  - Even though recommended by the EU[59], fully academic licenses (like MIT, BSD) on the other hand were not desirable for all partners that potentially would like to further exploit e.g. potential point-cloud compression algorithms, feature detection mechanism etc. later on without being able to control the spread of the underlying foreground developed in the DURAARK context
- **Datasets:** In the DURAARK project two main types of datasets can be distinguished by their provenance:

[55] Creative Commons – http://creativecommons.org/licenses/by-nd/3.0/legalcode
[56] OSI – http://opensource.org/
[57] LGLP – http://www.gnu.org/copyleft/lesser.html
[58] MIT License – http://en.wikipedia.org/wiki/GNU_Lesser_General_Public_License
[59] https://www.iprhelpdesk.eu/node/1901

- **Datasets acquired** from third parties such as industry stakeholders, public institutions like municipalities, publicly available resources or students at academic institutions. In the DURAARK case these are mainly:
  * Building Information Models in the Industry Foundation Classes format contributed by architectural offices or building owners
  * Point-cloud datasets produced by laser scan measurements by surveying companies
  * Linked Data vocabularies and datasets used for the semantic enrichment of preserved models or metadata. Examples include `DBPedia`, `buildingSMART bsDD`, `Getty AAT` and others described in WP 3

  Such datasets made accessible to DURAARK by contributors, are used to conduct part of the activities of the project, e.g. experimentation, model evaluation and validation of algorithms and concepts. Such datasets are protected by their original author and subject to licenses that might restrict redistribution. DURAARK will observe the licensing terms and abide to the terms of use. Where possible, publicly available datasets with permissive licensing models will be used to e.g. showcase IFC metadata extraction, point-cloud compression, point cloud registration to explicit IFC models, feature detection etc. This will ensure that the DURAARK results e.g. stemming for software prototypes can be reproduced, validated, compared and benchmarked to other approaches etc.

- **Datasets produced** by the DURAARK consortium within the context of the project. A prime example for this category of data are sets of metadata extracted, enriched and generated from IFC and E57 raw data during the archival ingestion. Such generated data will be provided in the Semantic Digital Archive reference implementation using the technical standards and best practises of the Linked Open Data community. It is made available to wide audiences and should be used by a wide range of stakeholders for experiments.

  Where these are descriptive and not violating ethical and privacy standards agreed to in the consortium agreement or are violating the IPR of the underlying data, the general approach of the DURAARK consortium is to make such produced data sets fully available under permissive open licenses. Other examples include mappings between vocabularies and datasets used for the semantic enrichment, registrations of point-cloud data sets, explicit geometry in the form of IFC files generated from real-world or simulated laser scans etc.

Especially for the latter category of *produced* datasets, the preferred license scheme is the Creative Commons license[60]. By default this will be a **Attribution** or **CC BY** or Open Data License **ODL**.

The Open Definition gives full details on the requirements for open data, knowledge and content. The full text of the license is available at Open Data Commons [61]

A more detailed assessment of IPR and the particular licensing involved can be found in deliverable D1.7 (IPR Management Plan).

5.4 Sustainability

The DURAARK consortium is working together to define a sustainability strategy for the project outcome after the project has ended. A number of key aspects have been identified that facilitate retaining and communicating the outcome to future stakeholders:

1. The outcome from DURAARK project must be known to the stakeholders
2. Outcome must be accessible and available
3. The organisation storing the results must be reliable[62]
4. The outcome has to be easy to access
5. Stakeholders need to be aware of how to use the outcome[63]
6. The outcome must be maintained – in the case of software, by managing the software and providing bug fixes etc.

Hence, one of the main activities of the *active* project is to kick-start the user base (the future stakeholders) by actively disseminating information about project results. From

[60] http://creativecommons.org
[61] Open Data Commons – http://opendatacommons.org/licenses/odbl/1.0/
[62] Cf. Trustworthy Digital Repository – http://www.trusteddigitalrepository.eu
[63] such as downloading source code from GitHub and building it locally

a kick-starter perspective we are distinguishing between two main types of audiences: academic and non-academic stakeholders, both of which we target differently. Depending on the type of audience, the dissemination activities so far have consisted of publishing academic articles (for academic stakeholders), presenting in conferences (where we meet both non-academic as well as academic stakeholders), and arranging targeted workshops for practitioners in the field (non-academic stakeholders).

Central pillars of dissemination and exploitation are the use of open licensing schemes and the wide publication of results through existing community platforms. These includes for instance dataset registries such as `DataHub`[64] or `DataCite`[65], software sharing communities and platforms (such as `GitHub`[66]) and the wide use of general social media and specialised community platforms. This takes advantage of the critical amount of individuals and organisations which can be reached and ensures a wide visibility of project outcomes.

In the longer perspective and after the conclusion of the project, we are currently looking into two orthogonal options; using a commercial preservation service (APA and APARSEN[67]) to ensure future access to reports and datasets, and joining forces with an organisation set to further preservation technology (Open Preservation Foundation[68] (OPF)).

The OPF is based on the organisation set in motion to manage the outcome of the PLANETS EU-project. The OPF has become a gravitational center for preservation-centric activities and they organise recurrent workshops. Additionally we are looking into utilising `COPTR` (Community Owned Digital Preservation Tool Registry) to make the software artefacts created in the course of DURAARK available to a larger audience. The latter may be combined with adding the software artefacts to `GitHub`[69] and Open Source-licensing the software.

Additionally, in order to maintain an engagement in DURAARK and its outcome, we are actively pursuing an effort to attract communities. A good example would be the W3C

[64] http://datahub.io
[65] http://datacite.org
[66] http://github.org
[67] http://www.alliancepermanentaccess.org
[68] http://www.openpreservation.org
[69] http://www.github.com

Linked Building Data group[70], where efforts are already invested. This work is ongoing and will continue through year 3.

By approaching an existing community and declaring a commitment to remain active within the community, we are in fact already expressing guarantees of presence for future stakeholders. Additionally creating a legal entity among the participants in DURAARK – which is orthogonal to community building – is likely to require both an administrative as well as financial burden.

5.5 Outcome centric sustainability and exploitation

5.5.1 WP2

Work package 2 is responsible for providing an integrated prototype for the functionality developed in the project. As described in Section 2.2 the integrated prototype is consisting of two parts which are in the responsibility of WP2 (the sustainability actions for the "Python Bindings" are described in Section 5.5.5):

- WorkbenchUI - the graphical user interface for consuming DURAARK's functionality from an end-user point of view
- Service Platform - a service-oriented platform allowing to consume DURAARK's functionality via an API for integration into custom applications

Our sustainability plan foresees three core points: Tooling, platform hosting and code maintenance.

5.5.1.1 Tooling

WP2 has a strong focus on preparing the code base and development environment to make it easily adoptable by third parties, e.g. (commercial) system integrators or research groups. With the complexity of the developed software system (currently 21 code repositories on Github which interact to constitute the WorkbenchUI and the Service Platform) it is not sufficient to only make the source code available for each sub-project and to document it. For the success and the community-acceptance of the project it is vital to provide third party developers with a state-of-the-art development environment

[70] http://www.w3.org/community/lbd/

to guarantee a low entry barrier when evaluating the project for their own purpose and of course for the code development itself afterwards. Furthermore, having a sound development environment in place also allows to maintain existing code more efficiently after project's lifetime. DURAARK's integrated prototype is developed as a web-application and a Web-API. The following list explains which state-of-the-art (and well-known in the community) web development services we are using:

Github[71] A collaborative source code management tool, issue tracker and documentation hub. See `http://github.com/DURAARK` for a listing of the code repositories.

Docker[72] A light-weight virtual machine implementation for deploying services and applications on servers. All our services and the WorkbenchUI come with a setup file for Docker.

Docker Hub[73] A registry for downloading pre-build docker images with a single click. All our services and the WorkbenchUI are distributed in this format for easy installation. See `https://registry.hub.docker.com/search?q=duraark` for a listing of our services on Docker Hub.

NPM Registry[74] A central registry for distributing software libraries. Helper libraries for bootstrapping a DURAARK service are located on NPM. See `https://www.npmjs.com/search?q=duraark` for a listing of the published services on NPM.

CircleCI[75] A continuous integration & deployment platform. All our code repositories use this service for running unit and integration tests before deploying new versions on our servers. See `https://circleci.com/gh/DURAARK/workbench-ui` for the continuous integration & deployment of the WorkbenchUI project. Replace the last part of the URL with the project name found on Github for viewing the other projects.

CodeClimate[76] An automatic code review service. All our Javascript-based services and the WorkbenchUI are using this service for monitoring the code quality. See `https://codeclimate.com/github/DURAARK/workbench-ui` for the code climate of the WorkbenchUI project. Replace the last part of the URL with the project name found on Github for viewing the other projects.

5.5.1.2 Platform hosting

FhA as work package leader is hosting a setup of the WorkbenchUI and the Service Platform on our infrastructure in Graz. The system is currently serving as a deployment and demo server. We plan to maintain this system beyond the course of the project for a minimum period of three years for a) showcasing the DURAARK project b) providing a live endpoint for the Service Platform so that interested third parties can try the API without having to install the system on their own servers and c) as deployment server for our additions to the project after its lifetime.

5.5.1.3 Code maintenance

The source code of the integrated software prototype is available on GitHub under the permissive MIT license. Each Github repositories contains sufficient documentation on how to use the software prototypes, or will contain the necessary documentation at the end of the projects lifetime. FhA has a strategic interest in maintaining and extending the Service Platform after the projects lifetime in the frame of running and future projects. Currently we are re-using the parts of the platform in the project Cultlab3D[77]. We are planning to use synergy effects with the project GINGER[78] (which calculates energy-efficiency algorithms on IFC models) in the future.

5.5.2 WP3

Partners involved in the activities in WP3 have a long standing interest in the sustained use and hence the maintenance of semantically rich Linked Data. The results of WP3, and in particular the prototypical software tools and collected data sets will be use beyond the lifecycle of the DURAARK project.

Next to common strategies such as the deposition of source code for the prototypes in publicly available repositories such as GitHub that will remain commonly accessible, the respective partners involved in WP3 also employ their independent sustainability strategies.

TUE has developed a DURAARK platform instantiation for the archival of Building

[77] http://www.cultlab3d.de/
[78] http://www.fraunhofer.at/en/visual-computing/projects/ginger.html

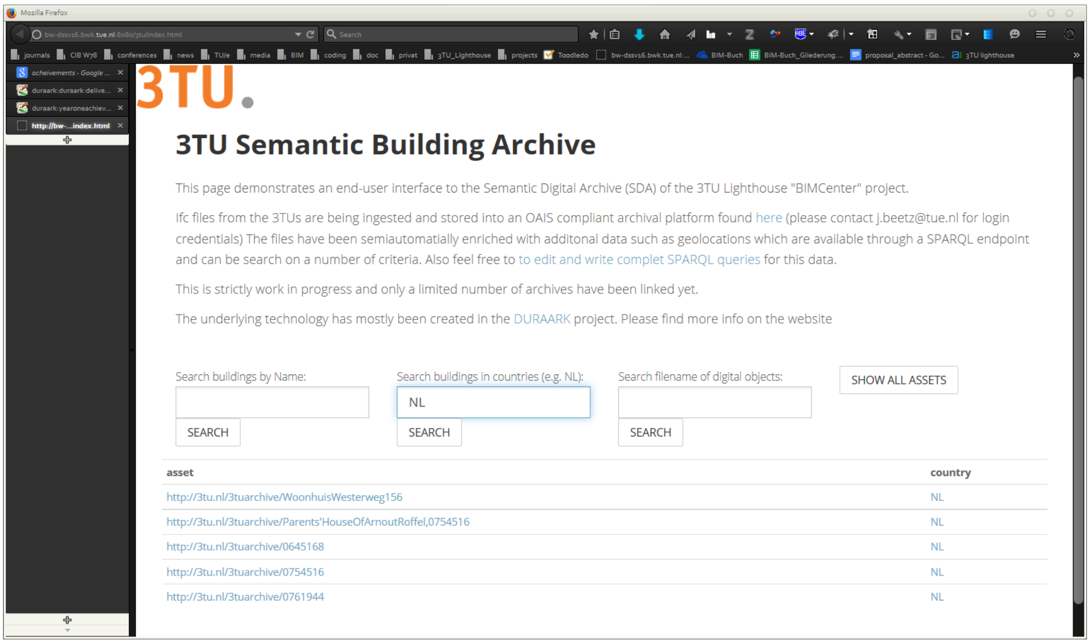


Figure 16: Screenshot of the query interface to the 3TU BIM archive for the sustained use of the DURAARK project results in academic environments.

Information Models produced in educational and research context at the three Dutch technical universities in Delft, Twente and Eindhoven. As part of a national research initiative "3TU Lighthouse projects" TUE secured additional funding for the implementation of a *3TU BIM Archive* that has been populated with an initial set of Building Information Models from partner universities in the Netherlands as a separate SDA instance. This SDA instance is also exposed by a public interface (see Figure 16 and Figure 17). Part of the funding have been used to secure a long-standing maintenance contract with the Cloud Hosting Platform Azure that will be maintained until 2018.

LUH is one of the main partners in WP3 and is responsible for a range of components as well as the datasets made available as part of SDA(S) and SDO. Components include the following:

- Focused Crawling component (deliverables D3.4 and D3.6)
- Semantic Enrichment component (entity linking) (deliverable D3.3)
- Clustering component (deliverables D3.4 and D3.6)

The source code of the aforementioned software components will be made available on

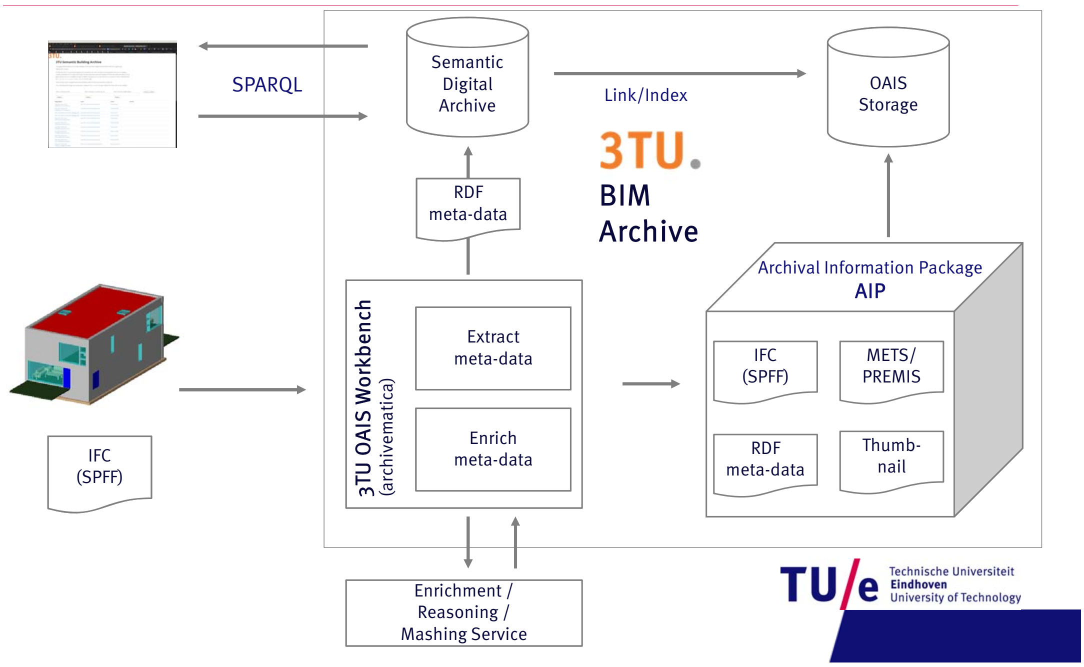


Figure 17: Schematic overview of the 3TU BIM Archive based on the DURAARK framework that will be in the funded, sustained use as a BIM archive of the three technical universities of the Netherlands.

GitHub and availability and continued maintenance will be carried for a period of at least three years. The datasets and endpoints – i.e. the data storage infrastructure and the actual data – associated with the SDA(S) and SDO will be sustained beyond the course of the project for a minimum period of three years. Licensing details for all of the above components are described in deliverable D1.7 (IPR Management Plan).

While all of the aforementioned tangible outcomes are part of ongoing research at L3S/LUH, it is intended to continue maintenance, development and expansion for an extended period of time. This will be facilitated, for instance, through reuse and further R&D work in other highly related projects, such as the ERC project ALEXANDRIA[79] or the European COST Action KEYSTONE[80], where key DURAARK project members are involved already. Further adoption by third parties and the potential for reuse, including in commercial settings, is particularly foreseen and supported.

[79] http://alexandria-project.eu/
[80] http://www.keystone-cost.eu/

5.5.3 WP4 & WP5

Work packages 4 and 5 are responsible for research and development of methods summarised by the term *geometric enrichment* and the subsequent delivery of software artefacts that can be seamlessly integrated into the DURAARK prototype, but can also be used in a stand-alone fashion or in other follow-up projects. The software components include:

- Difference Detection and Transfer of Semantics (deliverables D4.4.1, D4.2, and D4.3)
- IFCReconstruction and Compression (deliverables D5.5.1, D5.2, and D5.3)
- RISE (deliverables D5.2, D5.4, and D5.6)

The sustainability plan fo WP 4 and 5 comprises two components: Continuity of research and code maintenance.

5.5.3.1 Continuity of research

UBO has a long research tradition in areas and related methods that constitute the major ingredients for the DURAARK geometric enrichment component: Point cloud reconstruction, compression, and visualisation along with the application of sophisticated machine learning methods to geometric data are fundamental corner stones of its research program and will be further pursued after the project's end.

The developed methods along with the software artefact will thereby serve as an indispensable basis, which will ensure their continued development and expansion, especially in highly related research projects like Mapping on Demand[81] or HARVEST4D[82].

FhA also has strong interest in maintaining and further development of the RISE component for the detection and semantic description of nearly invisible shapes to continue research activities in the field of shape grammars started with projects like e.g. CITYFIT[83,84].

[81] http://www.ipb.uni-bonn.de/projects/MoD/index.php
[82] https://harvest4d.org/
[83] http://www.cgv.tugraz.at/CGV/Research/Projects/CITYFIT/
[84] Key personnel of the CITYFIT project is now working at FhA

5.5.3.2 Code maintenance

The before mentioned software components will be made available for a period of at least 3 years. Bugfixes and additional improvements that are added to the software during further development will also be made available as new releases.

5.5.4 WP6

Work package 6 is focusing on the long-term archival of architectural 3D data. The work package is lead by LUH / TIB – the German National Library of Science and Technology, who maintains an operational digital preservation system.

TIB is partaking in the project with the clear goal of extending functionality of its own digital preservation system towards the support of architectural 3D data. The short-term goal which is to be achieved within the project itself is therefore the inclusion of crucial digital preservation activities of file format identification and technical metadata extraction into the existing, OAIS-compliant digital preservation system.

With the integration of the file format identification pattern into DROID[85], the results are openly accessible.

File format extraction is supported by the e57 and IFC extraction tools presented in deliverable D2.4. The outputs are mapped to the technical metadata schemas `e57m` and `ifcm`, as presented in deliverables D3.1 and D6.2 and form the basis for the preservation planning and risk monitoring aspects pertaining to the objects' logical file format packaging.

The DURAARK extraction tools shall be added to TIB's digital preservation system. As the extractors are not tied to a specific digital preservation repository, the results can be integrated into any digital preservation workflow.

With TIB's preservation system being based on the Ex Libris Rosetta platform, the extractors and mappings will be rolled out to other Rosetta customers, e.g., through the

[85] DROID is developed and maintained by The National Archives of the United Kingdom (TNA). Any party can contribute to file format identification by submitting file format identification patterns to TNA's DROID Team.

Ex Libris Developer network[86]. As such, TIB will remain an active user of the file format identification and technical metadata extraction tools of the DURAARK project.

5.5.5 WP7

WP7 has developed means to integrate the developed DURAARK components in workflows of stakeholders, which are already existing or are expected to emerge in the near future. This approach visualises and exemplifies the DURAARK approach and presents tangible outcome to stakeholder communities.

A special effort was made to implement the DURAARK tools in land surveyor and architectural workflows in the software environments of the stakeholders 2.2.3. It is expected that the companies, that tested the tools so far will maintain and use the prototypical tools in the future.

The workflow implementation is developed in open formats and the open API of stakeholder software. A strong international community is here actively engaged in the development. The developed DURAARK workflow will engage through web platforms, as the Grasshopper3D website[87], with the respective communities. We currently have a dual strategy, sharing with both the `Revit Dynamo`[88] and the `Rhino Grasshopper3d`[89] communities.

The outreach to related communities through workshops and conferences with stakeholder organisations and other research units will be continued after the end of the DURAARK project by CITA. The past activities raised already strong interest in the DURAARK tools and approaches – here especially all tools that are related to the semantic enrichment of BIM and point-cloud data. It can be foreseen, that a strong demand will take place, when the DURAARK tools have achieved a certain degree of maturity and are all publicly available.

CITA is and will be using the so far developed workflows internally in related research projects, which continue after the DURAARK project, such as the 5 year Danish National

[86] https://developers.exlibrisgroup.com/rosetta
[87] http://www.grasshopper3d.com/
[88] http://dynamobim.com/forums/forum/dyn/
[89] http://www.grasshopper3d.com/

DURAARK DURABLE ARCHITECTURAL KNOWLEDGE

Elite research project ComplexModelling[90]. Another use of DURAARK approaches is going to take place in the Horizon2020 MarieCurie ITN Network InnoChain[91], where an ESR is investigating the integrated use of PointCloud scans in architectural design[92]. The DURAARK tools, will here be a good point of departure. CITA will finally maintain the collected DURAARK dataset on the DURAARK server for a three year period following the end of the project.

[90] http://www.complexmodelling.dk
[91] http://www.innochain.net
[92] http://innochain.net/innochain-phd-vacancies-2/#ESR5

6 Risk analysis

Each partner has a good overview of suitable dissemination activities in their domain. In contrast to other projects, DURAARK, however, has a high potential for involvement in standardisation processes. As touched upon in chapter 5.4, we are actively studying the feasibility of involving existing communities as a means to assume sponsorship of the output of DURAARK.

To actively engage in these processes, the right partners need to be identified and collaboration needs to be actively developed to achieve success.

6.1 Failing to make an impact on standardisation and establishment of defact standards

Risk: The consortium might have missed important partners and initiatives (collaborations) in order to generate the best impact on standardisation.

Risk assessment – Impact: High, **Probability:** Low

Description: Dissemination activities are planned according to the best opportunities which are identified at this stage.

Contingency solution: A plan for how the work is and will be carried out exists, but all available communication channels need to be monitored carefully to ensure that no crucial existing or new initiative – for standardisation or other collaborations - are missed. The state of the art and ongoing research developments will be monitored through ongoing clustering activities, via network activities and through regular attendance of scientific and industrial conferences. Furthermore, WP8 and the DURAARK coordination will keep a close contact with the DURAARK Advisory Board in order to ensure input is taken into account from all communities of relevance for DURAARK. Should new initiatives emerge which are so far unrecognised, the WP8 team will assess any collaboration opportunities and, if applicable, will plan new dissemination activities involving the new entities. The status of our efforts will be examined and, if needed, further plans adjusted during regular WP 8 meetings.

6.2 Failing to appropriately address all target communities

Risk: Given the high diversity of the project, there is a risk of missing out on important exploitation targets and communities.

Risk assessment – Impact: Medium, **Probability:** Medium

Description: A critical mass of users and take-up by key target audiences is crucial for the sustainability of the project. Given the high diversity of the project, addressing communities in a balanced way and reaching out to the right venues, working groups and communities is a constant challenge.

Contingency solution: As contingency actions, partners closest to so far under-represented community will be involved by WP8 leader LTU and targeted actions will be conducted. Constant monitoring of dissemination and exploitation action will help to alleviate and detect such issues early.

6.3 Biased stakeholder concern

Risk: Attention to different stakeholder groups gets out of balance (i.e. biased towards certain communities).

Risk assessment – Impact: Medium, **Probability:** Low

Description: The project has stakeholders in many areas which have to be reached through different activities at different times. While this involves a risk to under-recognise certain communities in favour of others, a certain focus might also emerge throughout the course of the project.

Contingency solution: While the DURAARK consortium involves partners from all key areas relevant to the project (digital preservation, building information modelling, semantic web), individual activities of partners are assumed to contribute to a balanced dissemination approach and will be complemented through additional dissemination actions. WP8 will permanently monitor dissemination activities and orchestrate joint dissemination activities which specifically target the identified dissemination needs.

6.4 Failing to successfully set-up a COST action

Risk: We fail to achieve a COST action concerning durable access to and use of building information.

Risk assessment – Impact: Low, **Probability:** Medium

Description: The consortiums exploitation and community-building strategy foresees the setup of a dedicated COST action, covering the inter-disciplinary topics of DURAARK. While COST Actions undergo a review and approval process, there is no guarantee, that even a well-prepared COST Action will finally received funding.

While we are not depending on this funding, it would to some extent affect the possibilities for post-project work and for expanding a community.

Contingency solution: The COST action is seen as an added pillar of the exploitation and sustainability strategy, which complements the overall set of activities. To this extent, it is not a mandatory element which can be complemented through other community-building activities.

7 Conclusions and Impact

This deliverable describes the initial version of the DURAARK market study and exploitation plan. While this document is going to be expanded in its final iteration (deliverable D8.7 – Market Study and Exploitation Plan v2), it already provides an overview of the *exploitation goals*, *strategy*, *actions* and the *sustainability strategy and commitments* for the project. While the project has produced a wide range of tangible outcomes, including actual *datasets* for DURAARK stakeholders in research and practice, a range of *software* prototypes and end-user applications, including an integrated DURAARK system ("DURAARK Workbench") and *reports and best practices*, ensuring the accessibility, reusability and usage is of utmost importance.

At a project-level, this collaboratively written document provides the general decisions and plans for implementing exploitation and sustainability actions as laid out in this deliverable. It is worth to note that deliverable D8.5 is complemented by additional, highly related documents, such as the DURAARK Consortium Agreement and the IPR Management Plan (deliverable D1.7), which further detail procedures and legal aspects involved in the exploitation of DURAARK foreground. Jointly with deliverable D8.5, these documents are of crucial importance for guiding exploitation and sustainability activities across the entire consortium and will be gradually refined throughout the upcoming reporting period.

Appendices

Appendix A

Detailed dataset listings

A.1 IFC files

In the following table, all available IFC datasets including their size in MB, IFC Schema version and availability are listed

Filename	Author/Owner	C/P	Type	[MB]	Schema	Tool
Academic_Barcelona-Pavillon_Arch_Non-conf.ifc	ADK	p	Pavillion	0,2	IFC2x3	Revit
Academic_DDB-Kingo_Arch_Non-conf.ifc	ADK	p	Single fam house	0,4	IFC2x3	Revit
Academic_DDB-KUA-context-massing_Arch_Non-conf.ifc	ADK	p	Context volumes	0,4	IFC2x3	Revit
Academic_DDB-massing_Arch_Non-conf.ifc	ADK	p	Context volumes	0,02	IFC2x3	Revit
Academic_DDB-Soeholm_Arch_Non-conf.ifc	ADK	p	Single fam house	3,2	IFC2x3	Revit
Academic_DDB-Uhrskov_Arch_Non-conf.ifc	ADK	p	Single fam house	1,2	IFC2x3	Revit
ATP_PR_Arch_Conf.ifc	Studio ATP	c	Factory	25	IFC2x3	Revit
Autodesk_Advanced-sample-project_Arch_Non-conf.ifc	ADK	p	Educational	37,2	IFC2x3	Revit
Autodesk_Basic-sample-project_Arch_Non-conf.ifc	ADK	p	Single fam house	5,3	IFC2x3	Revit
CCO_Lund-Kristallen_Arch_Conf.ifc	Studio CCO	c	Office	206	IFC2x3	Revit
DTU_Building-204_FM_Conf.ifc	-	c	Educational	37,8	IFC2x3	Revit 2014
KEJD_Haveje_Arch_Conf.ifc	KEJD	c	Kindergarden	51	IFC2x3	Revit
KEJD_Kastelsvej_FM_Conf.ifc	KEJD	c	Residential?	24,4	IFC2x3	Revit 2014
KIT_FJK_Arch_Non-conf.ifc	KIT	p	Single fam house	14	IFC2x2	ADT
KIT_Institute_Arch_Non-conf.ifc	KIT	p	Office	3	IFC2x3	ArchiCad 11
KIT_Smiley-West_Arch_Non-conf.ifc	KIT	p	Housing complex	3	IFC2x2	ArchiCad 11

continues on next page...

Table A.1 – continued from previous page

Filename	Author/Owner	C/P	Type	[MB]	Schema	Tool
LandGaard_Building-1_CONF.ifc		c				
LandGaard_Building-2_CONF.ifc		c				
LandGaard_Building-3_CONF.ifc		c				
LandGaard_Tower_CONF.ifc		c				
NBS_LakesideAC01_Arch_Conf.ifc	NBS	c	Restaurant	41	IFC2x3	ArchiCad
NBS_LakesideAC10_Arch_Conf.ifc	NBS	c	Restaurant	39	IFC2x3	ArchiCad
NBS_LakesideRVT01_Arch_Conf.ifc	NBS	c	Restaurant	37	IFC2x3	Revit
NBS_LakesideRVT10_Arch_Conf.ifc	NBS	c	Restaurant	37	IFC2x3	Revit
NBS_LakesideVW00_Arch_Conf.ifc	NBS	c	Restaurant	9	IFC2x3	VectorWorks
NBU_Duplex-Apt-COBie_Arch_Non-conf_Design.ifc	NBU	p	Residential	0,52	IFC2x3	?
NBU_Duplex-Apt-COBie_Arch_Non-conf_Handover.ifc	NBU	p	Residential	0,55	IFC2x3	?
NBU_Duplex-Apt-COBie_Arch_Non-conf_ProductInstall.ifc	NBU	p	Residential	0,56	IFC2x3	?
NBU_Duplex-Apt-COBie_Arch_Non-conf_ProductSelect.ifc	NBU	p	Residential	0,54	IFC2x3	?
NBU_Duplex-Apt-COBie_Arch_Non-conf_Programming.ifc	NBU	p	Residential	0,04	IFC2x3	?
NBU_Duplex-Apt-Coordination_Arch_Non-conf.ifc	NBU	p	Residential	2,3	IFC2x3	Revit
NBU_Duplex-Apt-Coordination_Arch_Non-conf_Optimized.ifc	NBU	p	Residential	1,6	IFC2x3	Revit
NBU_Duplex-Apt-Coordination_Eng-MEP_Non-conf.ifc	NBU	p	Residential	17,7	IFC2x3	Revit
NBU_Duplex-Apt-Coordination_Eng-MEP_Non-conf_Optimized.ifc	NBU	p	Residential	10,8	IFC2x3	Revit
NBU_Duplex-Apt-Sparkie_Eng-MEP_Non-conf_1.ifc	NBU	p	Residential	1,5	IFC2x3	Revit
NBU_Duplex-Apt-Sparkie_Eng-MEP_Non-conf_2.ifc	NBU	p	Residential	8,7	IFC2x3	Revit
NBU_Duplex-Apt-WSie_Eng-HVAC_Non-conf.ifc	NBU	p	Residential	31,3	IFC2x3	Revit
NBU_MedicalClinic_Arch_Non-conf.ifc	NBU	p	Clinic	17,7	IFC2x3	Revit
NBU_MedicalClinic_Arch_Non-conf_Optimized.ifc	NBU	p	Clinic	12,9	IFC2x3	Revit
NBU_MedicalClinic_Eng-CON_Non-conf.ifc	NBU	p	Clinic	18,9	IFC2x3	Revit
NBU_MedicalClinic_Eng-CON_Non-conf_Optimized.ifc	NBU	p	Clinic	18,9	IFC2x3	Revit
NBU_MedicalClinic_Eng-ELE_Non-conf.ifc	NBU	p	Clinic	11,7	IFC2x3	Revit
NBU_MedicalClinic_Eng-HVAC_Non-conf.ifc	NBU	p	Clinic	26,7	IFC2x3	Revit
NBU_MedicalClinic_Eng-MEP_Non-conf.ifc	NBU	p	Clinic	202,4	IFC2x3	Revit

continues on next page...

Table A.1 – continued from previous page

Filename	Author/Owner	C/P	Type	[MB]	Schema	Tool
NBU_MedicalClinic_Eng-MEP_Non-conf_Optimized.ifc	NBU	p	Clinic	122,8	IFC2x3	Revit
NBU_OfficeBuilding_Arch_Non-conf_1.ifc	NBU	p	Office building	4	IFC2x3	Revit
NBU_OfficeBuilding_Arch_Non-conf_2.ifc	NBU	p	Office building	0,6	IFC2x3	-
NBU_OfficeBuilding_Arch_Non-conf_Optimized.ifc	NBU	p	Office building	4	IFC2x3	Revit
NBU_OfficeBuilding_Eng-CON_Non-conf.ifc	NBU	p	Office building	10,8	IFC2x3	Revit
NBU_OfficeBuilding_Eng-CON_Non-conf_Optimized.ifc	NBU	p	Office building	10,8	IFC2x3	Revit
NBU_OfficeBuilding_Eng-ELE_Non-conf.ifc	NBU	p	Office building	6,4	IFC2x3	Revit
NBU_OfficeBuilding_Eng-HVAC_Non-conf.ifc	NBU	p	Office building	64,3	IFC2x3	Revit
NBU_OfficeBuilding_Eng-HVAC_Non-conf_Optimized.ifc	NBU	p	Office building	40,9	IFC2x3	Revit
NBU_OfficeBuilding_Eng-MEP_Non-conf.ifc	NBU	p	Office building	4	IFC2x3	-
NCC_Carlsby-detailed_Arch_Conf	NCC	c	mixed use	72	IFC2x3	
NCC_Carlsby-tender_Arch_Conf	NCC	c	mixed use	27	IFC2x3	
NVW_DCR-LOD100_Arch_Non-conf.ifc	NVW	p	Office building	0,2	IFC2x3	VectorWorks
NVW_DCR-LOD200_Arch_Non-conf.ifc	NVW	p	Office building	24,7	IFC2x3	VectorWorks
NVW_DCR-LOD200_Eng-CON_Non-conf_1.ifc	NVW	p	Office building	1,4	IFC2x3	VectorWorks
NVW_DCR-LOD200_Eng-CON_Non-conf_2.ifc	NVW	p	Office building	9,1	IFC2x3	Scia Engineer
NVW_DCR-LOD200_Eng-HVAC_Non-conf.ifc	NVW	p	Office building	32,4	IFC2x3	DDS-CAD
NVW_DCR-LOD300_Arch_Non-conf.ifc	NVW	p	Office building	281,3	IFC2x3	VectorWorks
NVW_DCR-LOD300_Eng-CON_Non-conf.ifc	NVW	p	Office building	10,8	IFC2x3	Scia Engineer
NVW_DCR-LOD300_Eng-HVAC_Non-conf.ifc	NVW	p	Office building	23,5	IFC2x3	DDS-CAD
Plan3D_Haus30_CONF.ifc		c				
Plan3D_Haus34_CONF.ifc		c				
PLH_DSV_Arch_Conf.ifc	PLH	c	Office	129	IFC2x3	Microstation
SGD_BARD_Arch_Non-conf.ifc	SGD	p	-	11,2	IFC2x2	Rom/Plantegning
SGD_BARD_Merged_Non-conf.ifc	SGD	p	-	42,7	IFC2x3	DDS-CAD
SGD_Blueberry_Arch_Non-conf_1.ifc	SGD	p	Single fam house	3,9	IFC2x3	AutoCAD Architecture 2009
SGD_Blueberry_Arch_Non-conf_2.ifc	SGD	p	Single fam house	5,6	IFC2x3	DDS-CAD Arkitekt

continues on next page...

Table A.1 – continued from previous page

Filename	Author/Owner	C/P	Type	[MB]	Schema	Tool
SGD_Blueberry_Arch_Non-conf_3.ifc	SGD	p	Single fam house	7,8	IFC2x3	DDS-CAD Construction
SGD_Blueberry_Eng-HVAC_Non-conf.ifc	SGD	p	Single fam house	10,4	IFC2x3	DDS-CAD Elektro
SGD_Blueberry_Eng-HVAC_Non-conf_Plumbing1.ifc	SGD	p	Single fam house	11,3	IFC2x3	DDS-CAD Elektro
SGD_Blueberry_Eng-HVAC_Non-conf_Plumbing2.ifc	SGD	p	Single fam house	4,8	IFC2x3	DDS-CAD Elektro
SGD_Blueberry_Eng-HVAC_Non-conf_Ventilation.ifc	SGD	p	Single fam house	8,9	IFC2x3	DDS-CAD Elektro
SGD_Blueberry_Merged_Non-conf_Plumbing1.ifc	SGD	p	Single fam house	11,3	IFC2x3	DDS-CAD Construction
SGD_Blueberry_Merged_Non-conf_Plumbing2.ifc	SGD	p	Single fam house	12,8	IFC2x3	DDS-CAD Construction
SGD_Blueberry_Merged_Non-conf_Ventilation.ifc	SGD	p	Single fam house	8,9	IFC2x3	DDS-CAD Construction
SGD_BODO_Arch_Non-conf_1.ifc	SGD	p	College	23,3	IFC2x2	ArchiCAD 7.0
SGD_BODO_Arch_Non-conf_2.ifc	SGD	p	College	23,3	IFC2x2	ArchiCAD 7.0
SGD_BODO_Arch_Non-conf_3.ifc	SGD	p	College	62,6	IFC2x3	ArchiCAD 10.0
SGD_BODO_Eng-HVAC_Non-conf_Plumbing.ifc	SGD	p	College	71,1	IFC2x2	PipeWork 6.35
SGD_BODO_Eng-HVAC_Non-conf_Ventilation.ifc	SGD	p	College	46,1	IFC2x2	DuctWork 6.35
SGD_CDB-2010_Eng-HVAC_Non-conf.ifc	SGD	p	-	2,9	IFC2x3	DDS-CAD
SGD_DDS-CAD_Eng-CON_Non-conf.ifc	SGD	p	Wall(?)	0,2	IFC2x3	Hjelpeprogram
SGD_Duplex_Eng-HVAC_Non-conf_Plumbing.ifc	SGD	p	Residential	90,6	IFC2x3	DDS-CAD
SGD_Duplex_Eng-HVAC_Non-conf_Ventilation.ifc	SGD	p	Residential	3,2	IFC2x3	DDS-CAD
SGD_HIBO_Arch_Non-conf.ifc	SGD	p	-	66,9	IFC2x3	ArchiCAD 11.0
SGD_HiTOS_Arch_Non-conf.ifc	SGD	p	-	51,6	IFC2x2	ArchiCAD 10.0
SGD_HiTOS_Eng-ELE_Non-conf.ifc	SGD	p	-	7,3	IFC2x3	DDS-CAD
SGD_HiTOS_Eng-HVAC_Non-conf.ifc	SGD	p	-	18	IFC2x3	DDS-CAD
SGD_Munkerud_Arch_Non-conf_1.ifc	SGD	p	Residential	2,8	IFC2x2	ArchiCAD 7.0
SGD_Munkerud_Arch_Non-conf_2.ifc	SGD	p	Residential	3,9	IFC2x2	ArchiCAD 8.0

continues on next page...

Table A.1 – continued from previous page

Filename	Author/Owner	C/P	Type	[MB]	Schema	Tool
SGD_Munkerud_Arch_Non-conf_3.ifc	SGD	p	Residential	2,9	IFC2x2	ArchiCAD 7.0
SGD_Munkerud_Eng-ELE_Non-conf.ifc	SGD	p	Residential	14,7	IFC2x2	Electrical Partner 6.35
SGD_Munkerud_Eng-HVAC_Non-conf_Sanitary.ifc	SGD	p	Residential	10,5	IFC2x2	PipeWork 6.35
SGD_Munkerud_Eng-HVAC_Non-conf_Ventilation.ifc	SGD	p	Residential	5,3	IFC2x2	DuctWork 6.35
SGD_Munkerud_Merged_Non-conf.ifc	SGD	p	Residential	34,4	IFC2x2	Electrical Partner 6.35
0573892_Tim-Dijkmans.ifc	TUE	p		7,3	IFC2x3	Revit
0588459_Florian-Roumen.ifc	TUE	p		0,6	IFC2x3	Revit
0593910_Emiel-Visser.ifc	TUE	p		0,6	IFC2x3	Revit
0613177_Ruben-Otten.ifc	TUE	p		1,2	IFC2x3	Revit
0626189_Veronika-Opletalova.ifc	TUE	p		0,9	IFC2x3	Revit
0645129_Burak-Kaan-Yilmazsoy.ifc	TUE	p		2,0	IFC2x3	Revit
0645168_Fatma-ceyda-gulserin.ifc	TUE	p		0,6	IFC2x3	Revit
0645195_Yuvacan-Atmaca.ifc	TUE	p		5,2	IFC2x3	Revit
0645207_Ayten-Basdemir.ifc	TUE	p		1,0	IFC2x3	Revit
0645548_Andres-Bravo.ifc	TUE	p		0,7	IFC2x3	Revit
0645658_Laura-Stefan.ifc	TUE	p		7,3	IFC2x3	Revit
0645666_Julien-Coutineau.ifc	TUE	p		0,3	IFC2x3	Revit
0647780_Alejandra-Garcia-Hooghuis.ifc	TUE	p		2,0	IFC2x3	Revit
0647827_c.b.jorgensen.ifc	TUE	p		0,6	IFC2x3	Revit
0648192_Elif-Aksayan.ifc	TUE	p		2,1	IFC2x3	Revit
0648296_aleksandra-murawska.ifc	TUE	p		2,8	IFC2x3	Revit
0648337_Anna-Kozlowska.ifc	TUE	p		0,3	IFC2x3	Revit
0648359_Emilia-Serowiec.ifc	TUE	p		0,3	IFC2x3	Revit
0648413_Harun-Cagatay.ifc	TUE	p		3,1	IFC2x3	Revit
0648435_Kalina-Kowalska.ifc	TUE	p		1,6	IFC2x3	Revit
0648462_Cyril-Combes.ifc	TUE	p		1,0	IFC2x3	Revit
0648577_Onem_Ayaz.ifc	TUE	p		2,2	IFC2x3	Revit
0648879_Paulina-Kowalczyk.ifc	TUE	p		3,6	IFC2x3	Revit
0648967_Clara-Giura.ifc	TUE	p		8,0	IFC2x3	Revit
0662891_Pau-Boluda-Hernandez.ifc	TUE	p		3,2	IFC2x3	Revit
0662945_Joan-Gaudin.ifc	TUE	p		1,2	IFC2x3	Revit
0663367_Goncalo-Pereira.ifc	TUE	p		0,6	IFC2x3	Revit
0663548_Daniel-Emmel.ifc	TUE	p		5,1	IFC2x3	Revit
0663603_Alexandra-Filip.ifc	TUE	p		3,6	IFC2x3	Revit
0664597_Ann-Katrin-Klemme.ifc	TUE	p		4,9	IFC2x3	Revit
0664717_Karina-Petrick.ifc	TUE	p		2,2	IFC2x3	Revit
0664795_Clara-Balta.ifc	TUE	p		3,7	IFC2x3	Revit
0664915_Stefano-Ortenzi.ifc	TUE	p		4,9	IFC2x3	Revit
0664959_Maria-Clemente.ifc	TUE	p		10,0	IFC2x3	Revit
0668471_Stan-Kruijssen.ifc	TUE	p		1,8	IFC2x3	Revit
0670665_Fabienne-Chene.ifc	TUE	p		1,9	IFC2x3	Revit
0672748_Anne-Lepesant.ifc	TUE	p		6,0	IFC2x3	Revit

continues on next page...

Table A.1 – continued from previous page

Filename	Author/Owner	C/P	Type	[MB]	Schema	Tool
0672962_Marketa-Brezinova.ifc	TUE	p		1,3	IFC2x3	Revit
0673641_Monica-Marteaux.ifc	TUE	p		2,3	IFC2x3	Revit
07036207_Jason-Chung.ifc	TUE	p		4,5	IFC2x3	Revit
0721677_Lisa-Fathalla.ifc	TUE	p		2,6	IFC2x3	Revit
0721771_Anna-Lloret-Cabot.ifc	TUE	p		1,4	IFC2x3	Revit
0723267_Sara Smibacker.ifc	TUE	p		0,4	IFC2x3	Revit
0723369_Francois-Loty.ifc	TUE	p		1,0	IFC2x3	Revit
0723386_Wesley-Lafage.ifc	TUE	p		0,3	IFC2x3	Revit
0723420_B.J.GELEY.ifc	TUE	p		8,7	IFC2x3	Revit
0723874_Amandine-Eyssartier.ifc	TUE	p		4,6	IFC2x3	Revit
0723948_Fantinel-Adelie.ifc	TUE	p		1,3	IFC2x3	Revit
0724193_Joana-Duarte.ifc	TUE	p		0,4	IFC2x3	Revit
0724489_Gulben-Basygit.ifc	TUE	p		2,5	IFC2x3	Revit
0724556_Sara-Battezzati.ifc	TUE	p		3,7	IFC2x3	Revit
0724599_Sara-Fernandez-Aguero.ifc	TUE	p		0,3	IFC2x3	Revit
0726863_Clara-Hernandez.ifc	TUE	p		1,2	IFC2x3	Revit
0726906_Senturk_Aysun_.ifc	TUE	p		2,1	IFC2x3	Revit
0728812_C.C.-Erol.ifc	TUE	p		13,4	IFC2x3	Revit
0729213_Gulsah-Kahraman.ifc	TUE	p		4,6	IFC2x3	Revit
0736210_Anna-Simina-Simaki.ifc	TUE	p		7,5	IFC2x3	Revit
0736224_Martin Slavicek.ifc	TUE	p		1,4	IFC2x3	Revit
0736238_Harold-Guy-Hanson.ifc	TUE	p		1,9	IFC2x3	Revit
0736245_Futura-Falistocco.ifc	TUE	p		3,6	IFC2x3	Revit
0736276_Margherita-Guzzoni.ifc	TUE	p		2,6	IFC2x3	Revit
0736289_Lamberti-Annachiara.ifc	TUE	p		1,9	IFC2x3	Revit
0738902_O-Sauer.ifc	TUE	p		0,3	IFC2x3	Revit
0739835_Roman-van-Os.ifc	TUE	p		0,7	IFC2x3	Revit
0745463_A.C.H.v.d.Laan.ifc	TUE	p		1,3	IFC2x3	Revit
0746284_Paulina-Krigsholm.ifc	TUE	p		4,0	IFC2x3	Revit
0746474_Karol-Kielczynski.ifc	TUE	p		3,4	IFC2x3	Revit
0747349_Tomas-Durdis.ifc	TUE	p		4,3	IFC2x3	Revit
0749114_Doruk-Can-Ozcifci.ifc	TUE	p		0,5	IFC2x3	Revit
0749986_Eren-Deniz.ifc	TUE	p		8,9	IFC2x3	Revit
0750440_frejaville-chloe.ifc	TUE	p		0,3	IFC2x3	Revit
0751352_Alejandro-Serra-Penalver.ifc	TUE	p		0,5	IFC2x3	Revit
0754516_Arnout-Roffel.ifc	TUE	p		0,6	IFC2x3	Revit
0759604_Laura-van-Loon.ifc	TUE	p		16,9	IFC2x3	Revit
0760714_Jochem-Rooswinkel.ifc	TUE	p		2,2	IFC2x3	Revit
0760746_Chris-Groenveld.ifc	TUE	p		4,2	IFC2x3	Revit
0761140_Stefan-Tack.ifc	TUE	p		0,7	IFC2x3	Revit
0761944_Nick-Rood.ifc	TUE	p		1,4	IFC2x3	Revit
0765153_Lester-Fei.ifc	TUE	p		4,2	IFC2x3	Revit
0782838_Laure-Pedot.ifc	TUE	p		0,3	IFC2x3	Revit
0784258_Emilia-Kierstan.ifc	TUE	p		6,2	IFC2x3	Revit
0784275_Laure-Pedot.ifc	TUE	p		8,2	IFC2x3	Revit
0784287_Philippe-Dibon.ifc	TUE	p		5,1	IFC2x3	Revit
0785246_Brambilla-Valentina.ifc	TUE	p		14,1	IFC2x3	Revit
0785328_Gabriel-Naudy.ifc	TUE	p		2,1	IFC2x3	Revit

continues on next page...

Table A.1 – continued from previous page

Filename	Author/Owner	C/P	Type	[MB]	Schema	Tool
0785516_Samineh-Alizadeh-Ashrafi.ifc	TUE	p		0,7	IFC2x3	Revit
0786416_Bagheri-Mahsa.ifc	TUE	p		2,0	IFC2x3	Revit
0786747_Salvagniac-Dorian.ifc	TUE	p		3,5	IFC2x3	Revit
0786752_Denis-Simonis.ifc	TUE	p		0,6	IFC2x3	Revit
0787445_Auxent-Sylvain.ifc	TUE	p		2,1	IFC2x3	Revit
0787489_Justine-Rossillion.ifc	TUE	p		7,0	IFC2x3	Revit
0787742_Natalia-Mylonaki.ifc	TUE	p		3,2	IFC2x3	Revit
0788237_Fariban-Salari-Joveini.ifc	TUE	p		1,0	IFC2x3	Revit
0788293_Aris-Santarmos.ifc	TUE	p		3,7	IFC2x3	Revit
0789505_Maria-Joao-Tato.ifc	TUE	p		1,5	IFC2x3	Revit
0796232_Jana-Zaverkova.ifc	TUE	p		0,8	IFC2x3	Revit
0796477_Martina-Iezzi.ifc	TUE	p		12,1	IFC2x3	Revit
0796596_Rutt-Vaikmae.ifc	TUE	p		4,8	IFC2x3	Revit
0796679_Andreaa-Onu.ifc	TUE	p		0,3	IFC2x3	Revit
0796702_Roger-Lee.ifc	TUE	p		3,9	IFC2x3	Revit
0822716_Elif-Yilmaz.ifc	TUE	p		7,0	IFC2x3	Revit
0822929_F.C.UNAL.ifc	TUE	p		11,9	IFC2x3	Revit
0822953_Ali-Kemal-Terlemez.ifc	TUE	p		4,0	IFC2x3	Revit
0829407_Rafal_muz.ifc	TUE	p		5,5	IFC2x3	Revit
085948_Tasci-Baran.ifc	TUE	p		11,0	IFC2x3	Revit
0_burcu-arinc.ifc	TUE	p		1,7	IFC2x3	Revit
648427_Caner-Bilgin.ifc	TUE	p		5,7	IFC2x3	Revit
Adrien-Pechberty.ifc	TUE	p		11,4	IFC2x3	Revit
Ahmet-Yildizhan.ifc	TUE	p		3,2	IFC2x3	Revit
Ali-Afrasiabi.ifc	TUE	p		2,7	IFC2x3	Revit
Alice-Cantore.ifc	TUE	p		4,8	IFC2x3	Revit
Allard-Mytrille.ifc	TUE	p		13,3	IFC2x3	Revit
Ana-Oliveira.ifc	TUE	p		2,2	IFC2x3	Revit
Bernhard-Bangert.ifc	TUE	p		2,9	IFC2x3	Revit
D'Amato-Lucia.ifc	TUE	p		27,7	IFC2x3	Revit
Eliott-Penel.ifc	TUE	p		6,6	IFC2x3	Revit
Fantinel-Adelie.ifc	TUE	p		1,7	IFC2x3	Revit
files.csv	TUE	p		0,0	IFC2x3	Revit
Francois-Loty.ifc	TUE	p		1,1	IFC2x3	Revit
Julian-Dario-Martos.ifc	TUE	p		2,1	IFC2x3	Revit
N.J.P.-Nancy.ifc	TUE	p		0,3	IFC2x3	Revit
Orkan-Zeynel-Guzelci.ifc	TUE	p		1,7	IFC2x3	Revit
Pinar-Basak-Tongal.ifc	TUE	p		2,6	IFC2x3	Revit
Sebastian-Forss.ifc	TUE	p		3,2	IFC2x3	Revit
W.-Schiermeyer.ifc	TUE	p		1,2	IFC2x3	Revit
0573892_Tim-Dijkmans.ifc	TUE	p		7,3	IFC2x3	Revit
0588459_Florian-Roumen.ifc	TUE	p		0,6	IFC2x3	Revit
0593910_Emiel-Visser.ifc	TUE	p		0,6	IFC2x3	Revit
0613177_Ruben-Otten.ifc	TUE	p		1,2	IFC2x3	Revit
0626189_Veronika-Opletalova.ifc	TUE	p		0,9	IFC2x3	Revit
0645129_Burak-Kaan-Yilmazsoy.ifc	TUE	p		2,0	IFC2x3	Revit
0645168_Fatma-ceyda-gulserin.ifc	TUE	p		0,6	IFC2x3	Revit
0645195_Yuvacan-Atmaca.ifc	TUE	p		5,2	IFC2x3	Revit
0645207_Ayten-Basdemir.ifc	TUE	p		1,0	IFC2x3	Revit

continues on next page...

Table A.1 – continued from previous page

Filename	Author/Owner	C/P	Type	[MB]	Schema	Tool
0645548_Andres-Bravo.ifc	TUE	p		0,7	IFC2x3	Revit
0645658_Laura-Stefan.ifc	TUE	p		7,3	IFC2x3	Revit
0645666_Julien-Coutineau.ifc	TUE	p		0,3	IFC2x3	Revit
0647780_Alejandra-Garcia-Hooghuis.ifc	TUE	p		2,0	IFC2x3	Revit
0647827_c.b.jorgensen.ifc	TUE	p		0,6	IFC2x3	Revit
0648192_Elif-Aksayan.ifc	TUE	p		2,1	IFC2x3	Revit
0648296_aleksandra-murawska.ifc	TUE	p		2,8	IFC2x3	Revit
0648337_Anna-Kozlowska.ifc	TUE	p		0,3	IFC2x3	Revit
0648359_Emilia-Serowiec.ifc	TUE	p		0,3	IFC2x3	Revit
0648413_Harun-Cagatay.ifc	TUE	p		3,1	IFC2x3	Revit
0648435_Kalina-Kowalska.ifc	TUE	p		1,6	IFC2x3	Revit
0648462_Cyril-Combes.ifc	TUE	p		1,0	IFC2x3	Revit
0648577_Onem_Ayaz.ifc	TUE	p		2,2	IFC2x3	Revit
0648879_Paulina-Kowalczyk.ifc	TUE	p		3,6	IFC2x3	Revit
0648967_Clara-Giura.ifc	TUE	p		8,0	IFC2x3	Revit
0662891_Pau-Boluda-Hernandez.ifc	TUE	p		3,2	IFC2x3	Revit
0662945_Joan-Gaudin.ifc	TUE	p		1,2	IFC2x3	Revit
0663367_Goncalo-Pereira.ifc	TUE	p		0,6	IFC2x3	Revit
0663548_Daniel-Emmel.ifc	TUE	p		5,1	IFC2x3	Revit
0663603_Alexandra-Filip.ifc	TUE	p		3,6	IFC2x3	Revit
0664597_Ann-Katrin-Klemme.ifc	TUE	p		4,9	IFC2x3	Revit
0664717_Karina-Petrick.ifc	TUE	p		2,2	IFC2x3	Revit
0664795_Clara-Balta.ifc	TUE	p		3,7	IFC2x3	Revit
0664915_Stefano-Ortenzi.ifc	TUE	p		4,9	IFC2x3	Revit
0664959_Maria-Clemente.ifc	TUE	p		10,0	IFC2x3	Revit
0668471_Stan-Kruijssen.ifc	TUE	p		1,8	IFC2x3	Revit
0670665_Fabienne-Chene.ifc	TUE	p		1,9	IFC2x3	Revit
0672748_Anne-Lepesant.ifc	TUE	p		6,0	IFC2x3	Revit
0672962_Marketa-Brezinova.ifc	TUE	p		1,3	IFC2x3	Revit
0673641_Monica-Marteaux.ifc	TUE	p		2,3	IFC2x3	Revit
07036207_Jason-Chung.ifc	TUE	p		4,5	IFC2x3	Revit
0721677_Lisa-Fathalla.ifc	TUE	p		2,6	IFC2x3	Revit
0721771_Anna-Lloret-Cabot.ifc	TUE	p		1,4	IFC2x3	Revit
0723267_Sara Smibacker.ifc	TUE	p		0,4	IFC2x3	Revit
0723369_Francois-Loty.ifc	TUE	p		1,0	IFC2x3	Revit
0723386_Wesley-Lafage.ifc	TUE	p		0,3	IFC2x3	Revit
0723420_B.J.GELEY.ifc	TUE	p		8,7	IFC2x3	Revit
0723874_Amandine-Eyssartier.ifc	TUE	p		4,6	IFC2x3	Revit
0723948_Fantinel-Adelie.ifc	TUE	p		1,3	IFC2x3	Revit
0724193_Joana-Duarte.ifc	TUE	p		0,4	IFC2x3	Revit
0724489_Gulben-Basygit.ifc	TUE	p		2,5	IFC2x3	Revit
0724556_Sara-Battezzati.ifc	TUE	p		3,7	IFC2x3	Revit
0724599_Sara-Fernandez-Aguero.ifc	TUE	p		0,3	IFC2x3	Revit
0726863_Clara-Hernandez.ifc	TUE	p		1,2	IFC2x3	Revit
0726906_Senturk_Aysun_.ifc	TUE	p		2,1	IFC2x3	Revit
0728812_C.C.-Erol.ifc	TUE	p		13,4	IFC2x3	Revit
0729213_Gulsah-Kahraman.ifc	TUE	p		4,6	IFC2x3	Revit
0736210_Anna-Simina-Simaki.ifc	TUE	p		7,5	IFC2x3	Revit
0736224_Martin Slavicek.ifc	TUE	p		1,4	IFC2x3	Revit

continues on next page...

Table A.1 – continued from previous page

Filename	Author/Owner	C/P	Type	[MB]	Schema	Tool
0736238_Harold-Guy-Hanson.ifc	TUE	p		1,9	IFC2x3	Revit
0736245_Futura-Falistocco.ifc	TUE	p		3,6	IFC2x3	Revit
0736276_Margherita-Guzzoni.ifc	TUE	p		2,6	IFC2x3	Revit
0736289_Lamberti-Annachiara.ifc	TUE	p		1,9	IFC2x3	Revit
0738902_O-Sauer.ifc	TUE	p		0,3	IFC2x3	Revit
0739835_Roman-van-Os.ifc	TUE	p		0,7	IFC2x3	Revit
0745463_A.C.H.v.d.Laan.ifc	TUE	p		1,3	IFC2x3	Revit
0746284_Paulina-Krigsholm.ifc	TUE	p		4,0	IFC2x3	Revit
0746474_Karol-Kielczynski.ifc	TUE	p		3,4	IFC2x3	Revit
0747349_Tomas-Durdis.ifc	TUE	p		4,3	IFC2x3	Revit
0749114_Doruk-Can-Ozcifci.ifc	TUE	p		0,5	IFC2x3	Revit
0749986_Eren-Deniz.ifc	TUE	p		8,9	IFC2x3	Revit
0750440_frejaville-chloe.ifc	TUE	p		0,3	IFC2x3	Revit
0751352_Alejandro-Serra-Penalver.ifc	TUE	p		0,5	IFC2x3	Revit
0754516_Arnout-Roffel.ifc	TUE	p		0,6	IFC2x3	Revit
0759604_Laura-van-Loon.ifc	TUE	p		16,9	IFC2x3	Revit
0760714_Jochem-Rooswinkel.ifc	TUE	p		2,2	IFC2x3	Revit
0760746_Chris-Groenveld.ifc	TUE	p		4,2	IFC2x3	Revit
0761140_Stefan-Tack.ifc	TUE	p		0,7	IFC2x3	Revit
0761944_Nick-Rood.ifc	TUE	p		1,4	IFC2x3	Revit
0765153_Lester-Fei.ifc	TUE	p		4,2	IFC2x3	Revit
0782838_Laure-Pedot.ifc	TUE	p		0,3	IFC2x3	Revit
0784258_Emilia-Kierstan.ifc	TUE	p		6,2	IFC2x3	Revit
0784275_Laure-Pedot.ifc	TUE	p		8,2	IFC2x3	Revit
0784287_Philippe-Dibon.ifc	TUE	p		5,1	IFC2x3	Revit
0785246_Brambilla-Valentina.ifc	TUE	p		14,1	IFC2x3	Revit
0785328_Gabriel-Naudy.ifc	TUE	p		2,1	IFC2x3	Revit
0785516_Samineh-Alizadeh-Ashrafi.ifc	TUE	p		0,7	IFC2x3	Revit
0786416_Bagheri-Mahsa.ifc	TUE	p		2,0	IFC2x3	Revit
0786747_Salvagniac-Dorian.ifc	TUE	p		3,5	IFC2x3	Revit
0786752_Denis-Simonis.ifc	TUE	p		0,6	IFC2x3	Revit
0787445_Auxent-Sylvain.ifc	TUE	p		2,1	IFC2x3	Revit
0787489_Justine-Rossillion.ifc	TUE	p		7,0	IFC2x3	Revit
0787742_Natalia-Mylonaki.ifc	TUE	p		3,2	IFC2x3	Revit
0788237_Fariban-Salari-Joveini.ifc	TUE	p		1,0	IFC2x3	Revit
0788293_Aris-Santarmos.ifc	TUE	p		3,7	IFC2x3	Revit
0789505_Maria-Joao-Tato.ifc	TUE	p		1,5	IFC2x3	Revit
0796232_Jana-Zaverkova.ifc	TUE	p		0,8	IFC2x3	Revit
0796477_Martina-Iezzi.ifc	TUE	p		12,1	IFC2x3	Revit
0796596_Rutt-Vaikmae.ifc	TUE	p		4,8	IFC2x3	Revit
0796679_Andreaa-Onu.ifc	TUE	p		0,3	IFC2x3	Revit
0796702_Roger-Lee.ifc	TUE	p		3,9	IFC2x3	Revit
0822716_Elif-Yilmaz.ifc	TUE	p		7,0	IFC2x3	Revit
0822929_F.C.UNAL.ifc	TUE	p		11,9	IFC2x3	Revit
0822953_Ali-Kemal-Terlemez.ifc	TUE	p		4,0	IFC2x3	Revit
0829407_Rafal_muz.ifc	TUE	p		5,5	IFC2x3	Revit
085948_Tasci-Baran.ifc	TUE	p		11,0	IFC2x3	Revit
0_burcu-arinc.ifc	TUE	p		1,7	IFC2x3	Revit

continues on next page...

Table A.1 – continued from previous page

Filename	Author/Owner	C/P	Type	[MB]	Schema	Tool
648427_Caner-Bilgin.ifc	TUE	p		5,7	IFC2x3	Revit
Adrien-Pechberty.ifc	TUE	p		11,4	IFC2x3	Revit
Ahmet-Yildizhan.ifc	TUE	p		3,2	IFC2x3	Revit
Ali-Afrasiabi.ifc	TUE	p		2,7	IFC2x3	Revit
Alice-Cantore.ifc	TUE	p		4,8	IFC2x3	Revit
Allard-Mytrille.ifc	TUE	p		13,3	IFC2x3	Revit
Ana-Oliveira.ifc	TUE	p		2,2	IFC2x3	Revit
Bernhard-Bangert.ifc	TUE	p		2,9	IFC2x3	Revit
D'Amato-Lucia.ifc	TUE	p		27,7	IFC2x3	Revit
Eliott-Penel.ifc	TUE	p		6,6	IFC2x3	Revit
Fantinel-Adelie.ifc	TUE	p		1,7	IFC2x3	Revit
Francois-Loty.ifc	TUE	p		1,1	IFC2x3	Revit
Julian-Dario-Martos.ifc	TUE	p		2,1	IFC2x3	Revit
N.J.P.-Nancy.ifc	TUE	p		0,3	IFC2x3	Revit
Orkan-Zeynel-Guzelci.ifc	TUE	p		1,7	IFC2x3	Revit
Pinar-Basak-Tongal.ifc	TUE	p		2,6	IFC2x3	Revit
Sebastian-Forss.ifc	TUE	p		3,2	IFC2x3	Revit
W.-Schiermeyer.ifc	TUE	p		1,2	IFC2x3	Revit

Table A.1: Detailed list of IFC datasets in use C/P: C=confidential P=public

A.2 Semantic Building Metadata - Linked Datasets

In the following table, an overview of some of the currently available data captured in the SDA(S)/SDO is provided. Please note that this is preliminary data, currently being significantly expanded. As part of the SDA, data has been generated generated by the different modules within DURAARK to produce an initial set of `buildM` instances, describing digital objects (e.g. IFC files) and physical objects (the actual buildings). From the initial process the data in the SDA is summarised by the following statistics. The generated resources are accessible through SDA's dereferencing mechanism, e.g. http://data.duraark.eu/resource/0721677.

#type	#resources	#triples
duraark:DigitalObject	74	730
duraark:PhysicalAsset	61	726
geonames:Feature	39	512
Total	174	1968

Table A.2: Initial `buildM` instances generated by DURAARK as part of the SDA.

While the SDO deals with creating snapshots of building-related Web data, initial crawls (see deliverables D3.2 and D3.3) have been produced covering specifically geodata and energy sustainability metadata (statistics below). More targeted **focused crawls** are currently under development as part of D3.6.

dataset	crawls	#namespaces	#types	#resources	#triples
enipedia	1	17	257	115623	1336758
clean-energy-data-reegle	1	2	8	775	11650
geological-survey-of-austria-thesaurus	1	4	8	9	75
Total	1	23	273	116407	1348483

Table A.3: Crawl statistics for DURAARK relevant datasets performed at 2014-07-06. The statistics show the number of performed crawls for the datasets, extracted namespaces and the corresponding resource types and instances.

In addition to this, the SDA also contains information about the dataset generated for our work regarding contextual semantic enrichment (presented in D3.3, D3.4). Here, we investigated the perception of architectural structures within the dataset (for more details regarding the architectural structures within the dataset, see http://data-observatory.

org/building-perception/. This data can be queried at the following Graph IRI in the SDAS: http://data.duraark.eu/structures/. Table A.4 presents some statistics regarding this data.

#type	#resources	#triples
Airports	100	1100
Bridges	59	600
Churches	139	1490
Halls	72	729
Skyscrapers	179	1800
Total	549	5719

Table A.4: Statistics regarding the perception of architectural structures (airports, bridges, churches, halls, skyscrapers) in USA and Germany.

A.3 E57 files

In the following table, all available E57 datasets including their size in MB, their number of scans and number of points are listed

Filename	Author/Owner	C/P	Type	[MB]	Scans	points$\times 10^6$
Acadia_Acadia_Office_NON-CONF.e57	Acadia	p	office	500	5	26,58
ATS_Restaurant_Restaurant-_CONF.e57	ATS	c	restaurant	7.500	13	352,79
Bentley_DescartesWorkshop-_academic_Conf.e57	Bentley	c	academic			
Bentley_PointoolsWorkshop-_academic_Conf.e57	Bentley	c	academic			
CITA_20131108-Room-24-7_NON-CONF.e57	CITA	p	room	1.000	1	41,78
CITA_DermoidI_Installation_NON-CONF.e57	CITA	p	installation	1.300	6	63,51
CITA_DermoidIII_NON-CONF.e57	CITA	p	installation	11.900	12	496,28
CITA_DesignHub_NON-CONF.e57	CITA	p	exhibition space	24.700	31	1021,95
CITA_IAAC_University_NON-CONF.e57	CITA	p	school			
CITA_IAAC-Menes_NON-CONF.e57	CITA	p	court yard			
CITA_Kronborg-CleanUp_NON-CONF.e57	CITA	p	castle	16.800	19	696,88
CITA_PersistentModelling02_NON-CONF.e57	CITA	p	installation	1.000	4	42,68
CITA_QualityOfScan-_ScannerSettings_005_Res1-5_Q2x-_NON-CONF.e57	CITA	p	room	600	1	26,77
CITA_QualityOfScan-_ScannerSettings_006_Res1-5_Q3x-_NON-CONF.e57	CITA	p	room	600	1	27,00
CITA_QualityOfScan-_ScannerSettings_007_Res1-5_Q4x-_NON-CONF.e57	CITA	p	room	600	1	27,09
CITA_QualityOfScan-_ScannerSettings_008_Res1-5_Q6x-_NON-CONF.e57	CITA	p	room	600	1	27,02
CITA_QualityOfScan-_ScannerSettings_009_Res1-20-_Q4x_NON-CONF.e57	CITA	p	room	30	1	1,56
CITA_QualityOfScan-_ScannerSettings_010_Res1-20-_Q6x_NON-CONF.e57	CITA	p	room	30	1	1,54
CITA_QualityOfScan-_ScannerSettings_011_Res1-20-_Q8x_NON-CONF.e57	CITA	p	room	30	1	1,53

continues on next page...

Table A.5 – continued from previous page

Filename	**Author/Owner**	**C/P**	**Type**	**[MB]**	**Scans**	**points**$\times 10^6$
CITA_QualityOfScan-_ScannerSettings_012_Res1-16-_Q3x_NON-CONF.e57	CITA	p	room	50	1	2,49
CITA_QualityOfScan-_ScannerSettings_013_Res1-10-_Q3x_NON-CONF.e57	CITA	p	room	140	1	26,77
CITA_QualityOfScan-_ScannerSettings_014_Res1-8_Q2x-_NON-CONF.e57	CITA	p	room	220	1	10,31
CITA_QualityOfScan-_ScannerSettings_015_Res1-5_Q2x-_NON-CONF.e57	CITA	p	room	670	1	26,77
CITA_QualityOfScan-_ScannerSettings_016_Res1-4_Q1x-_NON-CONF.e57	CITA	p	room	370	1	17,21
CITA_QualityOfScan-_ScannerSettings_017_Res1-2_Q1x-_NON-CONF.e57	CITA	p	room	1.500	1	70,26
CITA_QualityOfScan-_ScannerSettings_018_Res1-1_Q1-_NON-CONF.e57	CITA	p	room	3.400	1	158,84
CITA_ReferenceScans-_QualityOfPlanning_Better_NON-CONF.e57	CITA	p	room	2.500	3	128,11
CITA_ReferenceScans-_QualityOfPlanning_Poor_NON-CONF.e57	CITA	p	room	2.500	3	128,48
CITA_ReferenceScans-_QualityOfScan_Cleanness-_DarkPoints300_NON-CONF.e57	CITA	p	room	900	1	42,78
CITA_ReferenceScans-_QualityOfScan_Cleanness-_StandardFilters_DistanceBased-_NON-CONF.e57	CITA	p	room	900	1	41,44
CITA_ReferenceScans-_QualityOfScan_Cleanness-_StandardFilters_NON-CONF.e57	CITA	p	room	900	1	41,99
CITA_ReferenceScans-_QualityOfScan_Cleanness-_StrayPoints3_0-002_50_NON-CONF.e57	CITA	p	room	900	1	42,29
CITA_ReferenceScans-_QualityOfScan_Cleanness_Unclean-_NON-CONF.e57	CITA	p	room	900	1	43,52
CITA_ScanPrecision_000_Res1_5-Qua3x_NON-CONF.e57	CITA	p	room	640	1	26,42
CITA_ScanPrecision_001_Res1_4-Qua3x_NON-CONF.e57	CITA	p	room	1.000	1	41,53
CITA_ScanPrecision_002_Res1_2-Qua3x_NON-CONF.e57	CITA	p	room	4.000	1	168,21

continues on next page...

Table A.5 – continued from previous page

Filename	**Author/Owner**	**C/P**	**Type**	**[MB]**	**Scans**	**points**$\times 10^6$
CITA_ScanPrecision_003_Res1_4-Qua4x_NON-CONF.e57	CITA	p	room	1.000	1	41,74
CITA_ScanPrecision_004_Res1_4-Qua6x_NON-CONF.e57	CITA	p	room	1.000	1	41,64
LandGaard_Building-1_CONF.e57	LandGaard	c	residential			
LandGaard_Building-2_CONF.e57	LandGaard	c	residential			
LandGaard_Building-3_CONF.e57	LandGaard	c	residential			
LandGaard_Exterior_CONF.e57	LandGaard	c	residential			
LandGaard_Roof_CONF.e57	LandGaard	c	residential			
LandGaard_Tower_CONF.e57	LandGaard	c	residential			
LE34_132538-Norsgade-Gade-_CONF.e57	LE34	c	residential	3.000	5	142,86
LE34_132538-Norsgade-Norsgade-_CONF.e57	LE34	c	residential	37.000	62	1526,30
LE34_Gormsgade1_CONF.e57	LE34	c	office	7.000		
LE34_Gormsgade3-1sal_CONF.e57	LE34	c	office	3.000		
LE34_Gormsgade3-2sal_CONF.e57	LE34	c	office	1.500		
LE34_Gormsgade3-stue_CONF.e57	LE34	c	office	4.500		
LE34_Gormsgade3-ude_CONF.e57	LE34	c	office	2.000		
LE34_Kloak_CONF.e57	LE34	c	sewer	700		
LE34_Vestergade72-afleveret-1sal1-_CONF.e57	LE34	c	residential	30.000	52	1437,74
LE34_Vestergade72-afleveret-2sal-_bagtrappe1_CONF.e57	LE34	c	residential	2.000	4	110,24
LE34_Vestergade72-afleveret-2sal1-_CONF.e57	LE34	c	residential	30.000	51	1409,52
LE34_Vestergade72-afleveret-3sal2-_CONF.e57	LE34	c	residential	29.000	50	1382,98
LE34_Vestergade72-afleveret-4sal2-_CONF.e57	LE34	c	residential	22.000	40	1031,85
LE34_Vestergade72-afleveret-5sal2-_CONF.e57	LE34	c	residential	20.000	34	945,28
LE34_Vestergade72-afleveret-6sal2-_CONF.e57	LE34	c	residential	11.000	18	497,00
LE34_Vestergade72-afleveret-kaelder1_CONF.e57	LE34	c	residential	17.000	29	797,32
LE34_Vestergade72-afleveret-stue1-_CONF.e57	LE34	c	residential	21.000	36	998,15
Plan3D_Haus_34_Aussen_CONF.e57	PLAN3D	c	residential			
Plan3D_Haus_34_DG_CONF.e57	PLAN3D	c	residential			
Plan3D_Haus_34_EG1_CONF.e58	PLAN3D	c	residential			
Plan3D_Haus_34_EG2_CONF.e59	PLAN3D	c	residential			
Plan3D_Haus_34_EG3_CONF.e60	PLAN3D	c	residential			
Plan3D_Haus_34_GrosserHof1-_CONF.e61	PLAN3D	c	residential			
Plan3D_Haus_34_GrosserHof2-_CONF.e62	PLAN3D	c	residential			
Plan3D_Haus_34_GrosserHof3-_CONF.e63	PLAN3D	c	residential			
Plan3D_Haus_34_KG1_CONF.e64	PLAN3D	c	residential			

continues on next page...

Table A.5 – continued from previous page

Filename	Author/Owner	C/P	Type	[MB]	Scans	points×10^6
Plan3D_Haus_34_KG2_CONF.e65	PLAN3D	c	residential			
Plan3D_Haus_34_KG3_CONF.e66	PLAN3D	c	residential			
Plan3D_Haus30_Scene_Haus_130-_v2_CITAExport_Cluster-Aussen-_CONF.e57	PLAN3D	c	residential	28.000	49	1332,36
Plan3D_Haus30_Scene_Haus_130-_v2_CITAExport_Cluster-DG-_CONF.e57	PLAN3D	c	residential	10.500	18	496,98
Plan3D_Haus30_Scene_Haus_130-_v2_CITAExport_Cluster-Keller-_CONF.e57	PLAN3D	c	residential			
Plan3D_Haus30_Scene_Haus_130-_v2_CITAExport_Cluster-OG-_CONF.e57	PLAN3D	c	residential	23.000	40	1089,08

Table A.5: Detailed list of 57 datasets in use C/P: C=confidential P=public