



DURAARK
DURABLE
ARCHITECTURAL
KNOWLEDGE

D7.4 Evaluation Report

DURAARK

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

Date: 2016-01-20
Version 1.1
Document id. : duraark/2015/D.7.4/v1.1



Grant agreement number	:	600908
Project acronym	:	DURAARK
Project full title	:	Durable Architectural Knowledge
Project's 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 – Kunstakademiet 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.7.4/v1.1
Title of document	:	D7.4 Evaluation Report
Deliverable type	:	Report
Contractual date of delivery	:	2016-01-31
Actual date of delivery	:	2016-01-30
Lead beneficiary	:	CITA – Kunstakademiet Arkitektskole
Author(s)	:	Dag Fjeld Edvardsen < dag.fjeld.edvardsen@catenda.no > (Catenda), Jakob Beetz < j.beetz@tue.nl > (TUE), Henrik Leander Evers < heve@kadk.dk > (LTU), Thomas Krijnen < T.F.Krijnen@tue.nl > (TUE), Martin Hecher < martin.hecher@vc.fraunhofer.at > (FhA), Martin Tamke < martin.tamke@kadk.dk > (CITA), Mateusz Zwierzycki < mateusz.zwierzycki@gmail.com > (CITA), Michael Panitz < michael.panitz@tib.eu > (LUH), Raoul Wessel < wesselr@cs.uni-bonn.de > (UBO), Richard Vock < vock@cs.uni-bonn.de > (UBO), Sebastian Ochmann < ochmann@cs.uni-bonn.de > (UBO), Ujwal Gadiraju < gadiraju@l3s.de > (LUH)
Responsible editor(s)	:	Martin Tamke < martin.tamke@kadk.dk > (CITA),
Quality assessor(s)	:	Michael Panitz < michael.panitz@tib.eu > (LUH) Anders Lundkvist < anders.lundkvist@ltu.se > (Catenda)
Approval of this Deliverable	:	Stefan Dietze < dietze@L3S.de > (LUH) – Project Coordinator
Distribution	:	Public
Keywords list	:	DURAARK, evaluation, report

Executive summary

The Deliverable D7.4 documents and reports on the evaluation activities conducted by the DURAARK consortium. The evaluation addresses the diversity and interdisciplinarity of research and research output, that was conducted in DURAARK, which ranges from system and software components to processes and services.

The central question of all evaluation activities, was the fit to the current and future need of stakeholders. The evaluation activities focused on relevant target groups, which proved to have a vested interest in long-term archiving of data. Building upon the stakeholder related activities of previous years DURAARK was able to activate stakeholders to participate and provide input to the evaluation on all levels.

For this the whole consortium was involved in the evaluation activities and evaluated according to common evaluation criteria (Relevance, Utility - Integratability, Precision, Usability and Quality). These were differently weighted depending on the state of development, discipline and nature of the research outcome.

The activities necessitated the further development and adaptation of related methods in order to receive metrics, that allowed to evaluate the quality of the developed tools and techniques.

Taking into account, that DURAARK as a Specific Targeted Research Project (STREP) European Union Framework 7 project, does not aim to develop a product, the evaluation activities of the consortium show an overall good fit of the research outcome to the stakeholder expectations and needs. This includes all evaluated levels from the overall system and the Semantic and Geometric Enrichment components to the knowledge and practices developed in terms of Long Term Accessibility of 3D building data.

The evaluation activities were integral to the Year 3 of DURAARK and provided important feedback during the research in this period, as a solid ground and inspiration for the continued research after the DURAARK funding period.

Table of Contents

1	Introduction	9
2	Method	11
2.1	Approach	11
2.2	Targeted groups of stakeholders for the Evaluation	11
3	Risks	14
3.1	The group of stakeholders is not relevant or the selection is biased	14
3.2	The stakeholders do not understand the research presented to them	14
3.3	The group of stakeholder is too small	15
3.4	Failing to get in contact with the right stakeholder communities .	15
3.5	The software components developed in DURAARK are not mature enough to be evaluated	16
4	Overarching Evaluation Criteria	17
4.1	Relevance	17
4.2	Utility - Integratability	18
4.3	Precision	18
4.4	Usability	19
4.5	Quality	19
5	Evaluation on System Level	20
5.1	The DURAARK System - Deployability	20

	Aspects evaluated	20
	Method	21
	Results	21
5.2	Workbench UI - Functionality	24
	Aspects evaluated	24
	Method	24
	Datasets and Evaluation Activities	24
	Results	25
5.3	Service Platform – Integratability	28
	Integration into Stakeholder webplatform – BIMsync	28
	Integrating bimsync and DURAARK using the api from both	28
	Integration into Stakeholder work environment – Rhino3D	32
	Integration into Library environment - Metadata extractors (IFC, E57)	34
6	Evaluation - Semantic Enrichment Components	37
6.1	Enrichment (Focused Crawler)	37
	Aspects Evaluated	37
	Experimental Setup: Method and Datasets	37
	Results	42
6.2	IFC metadata extractor	44
	Aspects evaluated	44
	Method	45
	Datasets and Evaluation Activities	45
	Results	45
6.3	Entity Retrieval (SDAS)	49
	Aspects evaluated	49

Method	49
Datasets and Evaluation Activities	49
Results	50
Cluster Accuracy Evaluation	50
Entity Retrieval Evaluation	51
6.4 IFC - Point cloud schema	53
Aspects evaluated	53
Method	54
Datasets and Evaluation Activities	56
Results	57
6.5 IFC HDF5 Serialization	57
Aspects evaluated	57
Method	57
Datasets and Evaluation Activities	58
Results	58
7 Evaluation - Geometric Enrichment Components	60
7.1 Difference Detection	60
Difference Detection with DURAARK components	61
Difference Detection in Stakeholder Work Environment	63
Aspects evaluated	64
Method	64
Datasets and Evaluation Activities	65
Results	66
Summary	71
7.2 IFC reconstruction	73
Aspects evaluated	74

Method	74
Datasets and Evaluation Activities	76
Results	80
7.3 DURAARK components and Novel 3D Acquisition Techniques	94
Aspects evaluated	96
Method	96
Datasets and Evaluation Activities	96
Results	99
Conclusion	101
7.4 RISE	103
Aspects evaluated	103
Method, Datasets and Evaluation Activities	104
Results	106
Conclusion	109
8 Evaluation concerning Long Term Accessability	110
8.1 Long term archiving policies	110
Aspects evaluated	111
Method and Evaluation activities	111
Results	112
Conclusion	117
8.2 SIP Generator / Connection to Rosetta Ingest	118
Aspects evaluated	118
Method	118
Datasets and Evaluation Activities	119
Results	120
8.3 Metadata extractor (IFC, E57)	122

Aspects evaluated	122
Method	123
Datasets and Evaluation Activities	123
Results	123
9 Conclusion	124
1 Appendix	128
1.1 DURAARK WorkbenchUI Test - Contextualisation	129

1 Introduction

The overall usability of the DURAARK system and its software components, especially its ability to be adapted to different needs and environments of stakeholders, is evaluated and documented in this final evaluation report. This Deliverable D7.4 documents and reports on the evaluation activities conducted by the DURAARK consortium.

The evaluation activities have been carried out on all levels throughout the whole runtime of the DURAARK project.

- **Scientific Evaluation** carried out predominately through communities of peers in reviewed papers, journals, book chapters and presentations at scientific conferences.
- **Evaluation through Stakeholders** carried out through constant engagement and feedback from stakeholders in the project period through workshops, questionnaires, presentations and collaborations.

This Deliverable reflects both levels. A focus is however occurring on the stakeholder oriented evaluation activities, as much of the scientific evaluation is reported in the individual Deliverables of the other workpackages.

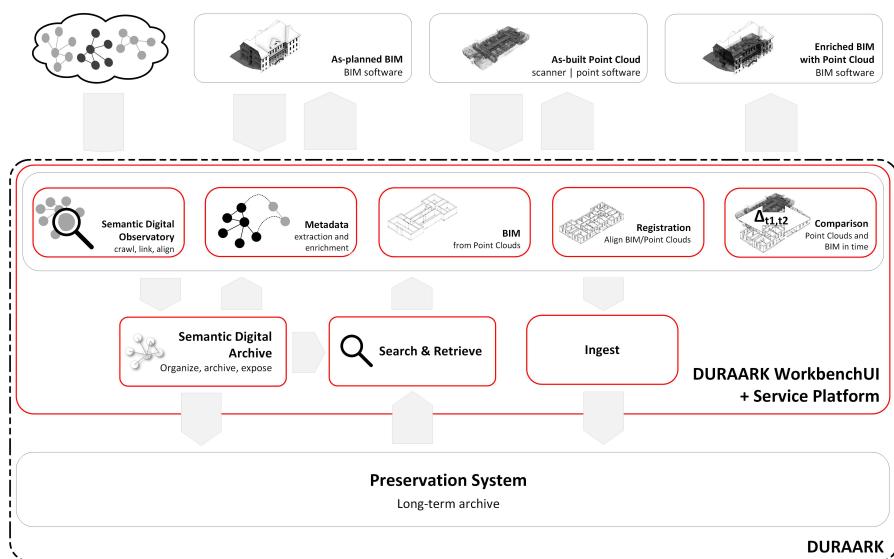


Figure 1: The DURAARK system is evaluated as a whole, as well as the main components, which ensure semantic and geometric enrichment of the ingested building data and the long-term accessibility of the data.

This document introduces in Section 2 the used evaluation methods and the general approach and stakeholders targeted in the related activities. After discussing the risks in Section 3, that are associated with this approach, we refine the in Deliverable D7.3 established evaluation criteria for the DURAARK research outcome in Section 4. The main body of this Deliverable is however the assessment of the DURAARK research outcome, differentiated according to the different aspects of the project in terms of:

- **System level** - Section 5
- **Semantic Enrichment level** - Section 6
- **Geometric Enrichment level** - Section 7
- **Long Term Accessability of Building Data** - Section 8

The document closes with an overarching assessment of the conducted evaluation activities and an outlook of future work in Section 9.

2 Method

The evaluation activities took into account, that DURAARK research and research output is of inherently diverse and interdisciplinary nature. It ranges from system and software components to processes and services. The central question of all evaluation, was the fit to the stakeholder's current and future needs.

Due to the heterogeneity of the research each evaluation activity had to employ a fitting evaluation method, with credential in the respective research community. The used method is hence described in the start of each section of this report, followed by an explanation of the datasets for and the results of the evaluation activities.

2.1 Approach

The general evaluation approach in this Deliverable is based on the evaluation criteria, which have been introduced in D7.3. These criteria where the base for all evaluation activities, which took place in year 3 of DURAARK.

Following the recommendation issued by the reviewers after the second review of the DURAARK project in Luxembourg on 25.-26.03.2015 a more precise definition of addresses target groups and the overarching evaluation criteria was set up. These are described in the following two sub-chapters. Evaluation methods specific to each research outcome of DURAARK, were defined during the discussions of the DURAARK partnership in the Workpackage 8 in year 3.

2.2 Targeted groups of stakeholders for the Evaluation

The Deliverable focuses in its evaluation activities on a set of target groups, which proved to have a vested interest in the overall goal of the DURAARK project: to create long-term access to building data.

The Deliverable D7.3 described the different stakeholders, which engage in the creation, modification, use and archival in the emerging digital lifecycle of building data. The result of this analysis is a set of stakeholders with such a long-term interest in building data. Only these stakeholders have a natural interest to archive building related data - whether in order to sustain the value of the data, because they invested their own capital (such as Institutional Building Owners) or because they are tasked to create and maintain systems

to preserve the value of the data (such as Cultural Heritage institutions and AEC Software vendors). The communalities between the stakeholders is, that they actively maintain and keep building data.

The Deliverable D7.3 created an important distinction between Data Users and Data Owners. Building Data is the property of **Data Owners**. These conduct typically data maintenance in a regular fashion. This might include the update of objects or the object's properties in building data, as it is for instance the case during Facility Management activities. **Data Users** show a distinctively different approach towards data. This group receives the data for a certain amount of time from the Data Owner in order to update and modify it. This can be i.e. through renovation, addition of buildings or even demolition and rebuilding of an architectural structure. Upon completion of these activities the Data Users hands an updated dataset back to the Data Owner. The emerging lifecycle of digital data between the stakeholders is captured in the Building Data Lifecycle Diagram (fig. 2, which was developed in Deliverable D7.3 [2](#)).

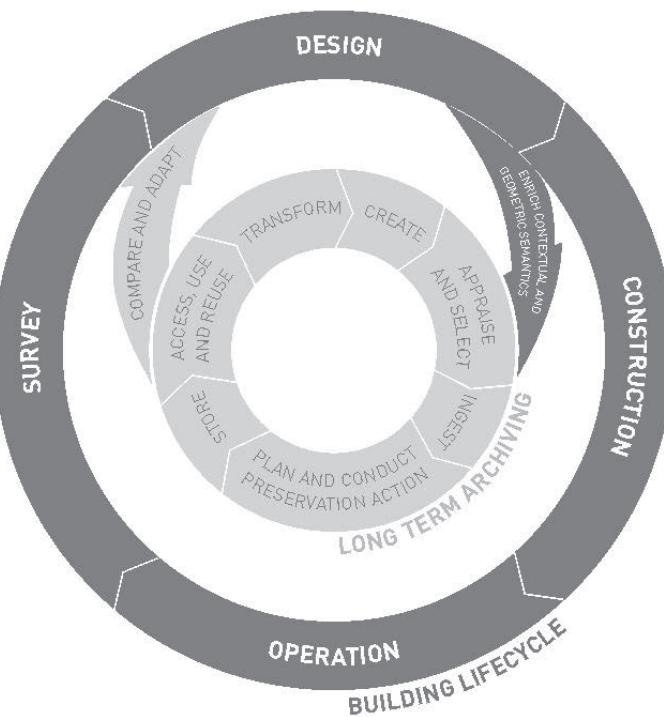


Figure 2: The emerging lifecycle of building data sees data flowing between the realm of archiving executed by the Data Owner and use of data for modifications and update of buildings through the Data User.

This Deliverable differentiates between the above described two major groups of users and declares for every evaluation activity, which stakeholders are addressed and activated.

- **Data owners** included stakeholder groups: Cultural heritage, Institutional Building Owners, AEC Software vendors
- **Data users** included stakeholder groups: Architect /Engineers, Land Surveyors, Institutional Building Owners

The direction and development of the DURAARK research is based on a set of evidence based predictions concerning the current and future use of building related data in the building data related professions. The investigation and constant update of the validity of these conditions under stakeholder specific aspects were part of the evaluation activities throughout the whole project period. Results were first laid out in the DURAARK Deliverable D7.1 and than followed up in D7.2, D7.3 and D8.5. The observations are backed up by observations of current developments in the field, as the creation of a *Linked Data Working Group* in the buildingSMART organisation ¹ or the ongoing efforts in the European Commissions INSPIRE directive ². The engagement of DURAARK in these activities and the contact with many stakeholder communities, as described in the other reports of WP 7 and 8, illustrate however, that progressive practices around building data are just emerging among stakeholders and that those with first hand knowledge on the use of i.e. semantic data, Point Clouds and data-driven design practice [3] are sparse. The DURAARK project managed nonetheless to engage with small, but well informed communities of stakeholders for the evaluation activities.

¹<http://www.buildingsmart.org/standards/standards-organization/groups/linked-data-working-group/>

²<http://inspire.ec.europa.eu/>

3 Risks

The evaluation activities have been carried out with a set of risks and contingencies actions in mind. These are documented here.

3.1 The group of stakeholders is not relevant or the selection is biased

Risk: The selection of stakeholders for the evaluation of DURAARK research outcome is not relevant or the selection is biased towards, i.e. non-critical stakeholders.

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

Description: The validity of evaluation activities for and with stakeholders depends on having a group which is representative of the stakeholders community under investigation.

Contingency solution: Especially Work Package 7 has been active to identify stakeholders, that are representative of a stakeholders communities work practice. as these practices are however just emerging in terms of information based building models or, as in the case of the cultural heritage community, do hardly exist at all at the current point of time, DURAARK selected candidates from the stakeholder communities, which demonstrate in terms of BIM progressive work practice.

3.2 The stakeholders do not understand the research presented to them

Risk: The stakeholders are not able to evaluate the research presented to them for evaluation or they evaluate it in the wrong way, as they are missing insights or context.

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

Description: As DURAARK is not aiming at the production of a solid system or single product, stakeholders are naturally executing their evaluation on somewhat isolated parts. These have to be contextualised, so that stakeholders get a sufficient understanding of the

tool and necessary processes, that would take part before and after the component.

Contingency solution:

3.3 The group of stakeholder is too small

Risk: The group of stakeholders is too small for an evaluation.

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

Description: The group of stakeholders is too small for the evaluation of a component with quantitative methods and to get representative answers.

Contingency solution: The risk can be addressed through efforts to increase the numbers or to switch the evaluation method. Quantitative methods can as well be supported by qualitative ones.

3.4 Failing to get in contact with the right stakeholder communities

Risk: Failing to get in touch with the right stakeholder communities of the evaluation of approaches, so that we are not able to evaluate towards their needs.

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

Description: Depending on target (academy, practitioners within the field, industry) different activities are necessary to identify and reach them - ranging from visit at fairs and conferences, web searches to meetings and video conferences. The DURAARK project is aware, that stakeholders have limited resources in terms of time to participate in non-profit research activities and that stakeholder engagement requires activity and persistence.

Contingency solution: This is an ongoing effort where each partner have to be active. Especially the work in WP7 has been to identify and build up resilient relations with stakeholder, that can be used for evaluation tasks. In the case, that these relations turn out to be not be good enough to perform evaluation tasks the WP7 leadership will get noticed and use the existing network in all stakeholder communities to identify alternative stakeholders for the tasks.

3.5 The software components developed in DURAARK are not mature enough to be evaluated

Risk: The software components developed in DURAARK are not mature enough to be evaluated.

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

Description: Research and development is notorious for delays. These delays might either shorten the period available for evaluation towards a threatening low amount or fail to produce working prototypes and components for evaluation.

Contingency solution: DURAARK employs agile development strategies. DURAARK has for instance introduced the use of Docker containers in order to reduce the amount of time necessary to cross operation systems for evaluation activities. The close contact and exchange between workpackages and within the partners of a workpackage provide furthermore constant feedback about the state of development. Evaluation actions can finally be adjusted to the state of component development, as long as the general evaluation criteria are adhered.

4 Overarching Evaluation Criteria

The DURAARK consortium agreed in its General assembly in Copenhagen in November 2014 a set of five overarching criteria for the evaluation activities. These were introduced and discussed in more detail in the Deliverable D7.3. The evaluation activities, which are reported on in this Deliverable are carried out in regards to this set of criteria listed below. Each of the chapters in the three major sections 5, 6 and 7, 8 of this Deliverable declare as well, which criteria they are evaluating towards.

4.1 Relevance

Is the research outcome of DURAARK of relevance for the practices of the building data related stakeholder communities?

These practices are in constant flux with the ongoing shift towards information based practices, as described in D7.3 in great detail for stakeholders, which engage with architectural data. The relevance of research output can hence not solely be evaluated against current practices, but has to be projected against scenarios of emerging and future practices. These have been drawn up in the Deliverable D7.1, D7.2 and D7.3, which point for instance out, that already now all work processes of the major stakeholder groups are centred around building information, rather than for instance 2D drawings. These stakeholders named BIM and 3D scan data as major contributors to the information, which they have about buildings.

The DURAARK project based the question of the relevance of the research for stakeholders on interviews with early adopters among the stakeholders, feedback from online communities during the runtime of the project and indicators, based on the observation of trends among relevant stakeholder groups. All of these show a general shift of the stakeholder practices from manual to automated processes - and a lack of related tools. Deliverable D7.3 reported on these shifts and provided evidence for it in several building data related processes among the investigated stakeholders, including enrichment, linkage, ingest or validation of archived building data.

An important step in the evaluation of the relevance of the DURAARK research is the identification and observation of stakeholders with novel and highly relevant practices

³ or questions ⁴). Their feedback to the DURAARK research served in the processes that led to this Deliverable as indicator, whether and where DURAARK processes are of relevance in future processes of stakeholders. We found, that the use cases developed in WP2 provided a sufficient base to describe and classify these.

4.2 Utility - Integratability

What is the effort to integrate the developed approaches and tools into stakeholder work environments?

All evaluation activities in DURAARK carried out towards this criteria have to be seen in the light of DURAARK being a STREP Project: the developed solutions are prototypical and did not see the same effort in terms of utility, as developed products with a clear vision for the target group can have.

The DURAARK solutions can however be evaluated in terms of interoperability and modularity. The utility of the semantic enrichment and cloud-based processes within the DURAARK project can be assessed through an evaluation of the used vocabularies connection to the building domain (i.e., whether they use already established standards such as the buildingSMART data dictionary and can be used as well on localized implementations and derivatives of these). This evaluation covers the archival actions necessary to ingest, as well as to retrieve and re-use building data in the building information lifecycle.

4.3 Precision

How precise are the devloped DURAARK components?

A qualitative evaluation of the precision of the developed DURAARK methods can be used for instance for the automated linkage of data, detection of architectural structures within point-clouds and their relation and transformation to BIM models ⁷. The DURAARK datasets have to suffice for these needs.

³e.g. the Swedish municipality of Falun, who has a huge repository of 3D scans, but little means to extract semantically rich data from these on architectural structures, which in return can be used by commissioned companies

⁴e.g. the Danish Bygningsstyrelsen that has to decide whether it wants to keep data archived in cloud based storages, when the underlying proprietary system will be updated

4.4 Usability

Do the developed DURAARK approaches fit into stakeholder workflows?

The assessment of the usability of the prototypes focuses on questions such as whether the developed approaches fit into existing workflows of users and into their changing technological environments 5.3. Questions of GUI and interaction with users are certainly as well of concern. A detailed evaluation on a qualitative level is however not within the scope of a STREP research project.

4.5 Quality

How good is the approach of components in comparison to the groundtruth provided by established approaches?

A qualitative assessment is used for the approaches developed in WP3, at the level of the Semantic Digital Archive / Observatory or the semantic data enrichment 6.

5 Evaluation on System Level

The *DURAARK System* is the overall term for describing the two elements of functionality which were developed in the course of the project:

DURAARK Service Platform A web-based, service-oriented platform providing functionality developed in the project via a REST-API.

DURAARK WorkbenchUI A graphical user interface to showcase the functionality developed in the project within a web-application. The web-application uses the DURAARK Service Platform's API in the background and serves as a reference implementation on how to use the platform in a graphical user interface.

The system level evaluation is focused on three central aspects:

DURAARK WorkbenchUI Functionality How well are the user requirements from D2.2.1 reflected in the graphical user interface?

DURAARK Service Platform Integrateability How flexible is the integration of functionality into stakeholder's workflows and applications?

Deployability How user-friendly is it to setup the system on a stakeholder's premise?

The following sections answer those three questions in detail and together give a holistic assessment of the usability of the DURAARK System. The section will also give recommendations for improvements based on the evaluation for future work on the system (see D8.7, for a detailed plan on how to sustain the development work done in the project).

5.1 The DURAARK System - Deployability

Aspects evaluated

This evaluation is a qualitative assessment of the deployment usability -of the DURAARK System. The goal of the evaluation was to see if the system is installable for non-technical experts (local installation), but also for IT administrators when deploying the system in a network environment to have a central instance of the system which users can collaboratively work on (server installation).

- Used operating system?

- Is the documentation sufficient for installing the system?
- Which problems did arise?
- Was it necessary to contact FhA for support?
- Are there enhancement suggestions?

Method

Local Deployment For the local deployment our target system was Microsoft Windows, as the majority of the target stakeholders is using this operating system on their local computers. All project partners were conducting a test installation of the DURAARK system on a Windows computer and were reporting the results to WP2. As a base for the installation the partners were provided with the publicly available installation instructions for Windows⁵. These installation instructions were written as a result of integrating the DURAARK system into the Rhino3D done by CITA (see 5.3).

Server Deployment <https://www.docker.com/docker-kitematic>

The server installation targets the Linux operating system, our tests were conducted on the distribution Ubuntu 14.04 LTS Server Edition (the installation is possible on all Linux distributions supporting the Docker technology). FhA provided installation instructions for Linux⁶, which were executed by Catenda to integrate the DURAARK System with their bimsync web application (see 5.3). The following results contain the experiences of Catenda and FhA when installing the system as a network-wide service on a Linux server.

Results

Local Installation For the local deployment the main software pre-requisite is the Docker Toolbox⁷. The Docker Toolbox is an effort of the Docker team to support their technology on Windows for non-technical users (Linux systems were the starting point of the technology). This effort is very young and only a few months old. The majority of problems when installing the DURAARK System on Windows were resulting from

⁵Windows installation documentation: <https://github.com/DURAARK/duraark-installer/wiki/Windows-Installation-Instructions>

⁶Linux Installation Instructions: <https://github.com/DURAARK/duraark-installer/wiki/Linux-Installation-Instructions>

⁷Docker Toolbox: <https://www.docker.com/docker-toolbox>

issues with the Docker Toolbox. The Docker Toolbox is evolving fast, meaning that its robustness will grow in the next months, as it has a large community behind the development efforts.

When the pre-requisite Docker Toolbox was installed the deployment of the DURAARK System is done through the duraark-installer⁸. The installer provides a "single click" experience for downloading and configuring the system. The deployment tests conducted by partners helped to identify two recurring issue: limited space and an issue with the configuration of the system at installation time. The space issue is caused by the Docker Toolbox, which configures a virtual machine on the Windows host with a space limit of 20GB. The DURAARK System is currently using that amount of space, causing the installation to abort silently. To mitigate that problem WP2 provides a pre-configured and correctly sized virtual machine with the system installed for stakeholders to download and run⁹. In future versions of the Docker Toolbox the size of the virtual machine will be configurable, which will resolve that issue. For the second issue, the mis-configuration of the DURAARK System, WP2 extended the installation instructions to contain a "Troubleshooting" section that helps to resolve this issue with a tool provided by the duraark-installer¹⁰.

CITA was the main partner during the development of the duraark-installer for Windows. This is a quote of their experiences with deployment of the DURAARK System:

"We installed and tested the duraark-installer on a 64bit Windows machine at CITA. We found the installation documentation sufficient for a non-technical user to install the system. With the installation of the actual Docker Toolbox we experienced, that it took quite a few steps and trials to get a stable running version on a machine. This does not stem from the duraark-installer though, but originates from the fact that Docker for Windows is a quite young package. This will get better as the Docker Toolbox matures; which is happening at a rapid pace at the moment. Besides this, we also experienced the individual Docker containers not to unload their memory consumption of the Virtual Machine. This was though fixed through a few iterations with WP2, which has resulted in the duraark-installer v0.4.1."

⁸duraark-installer: <https://github.com/DURAARK/duraark-installer/wiki>

⁹FIXME: provide download URL

¹⁰Troubleshooting: <https://github.com/DURAARK/duraark-installer/wiki/Windows-Installation-Instructions#troubleshooting>

Server deployment The server deployment test were conducted on Linux operating system (i.e., Ubuntu 14.04 LTS Server Edition). The pre-requisite of installing the DURAARK System is Docker. The installation of Docker on Linux is straightforward and well documented by the Docker team¹¹. The installation of the DURAARK System itself is done on the command line with the following commands:

```
> git clone https://github.com/DURAARK/duraark-system.git  
> cd duraark-system  
> sudo docker-compose up -d  
> sudo mkdir -p /duraark-storage/sessions
```

The only issue for partners testing the setup were caused by limited space on the test server, as the system needs about 20GB of space to work correctly. Other than that the deployment of the system on Linux hosts is working well.

Due to its robustness and deterministic behaviour WP2 recommends stakeholders to deploy the DURAARK System as a network-wide server deployment, which also allows for easy updates on a centralized place.

Catenda was the main partner to develop the deployment on Linux, here is a quote of their experiences: " We created a 64-bit Linux Virtual Machine (Ubuntu 14.04 LTS) with 4 processor cores, 4 Gigabytes of RAM and 50 Gigabytes of harddrive space. Then we followed the installation guidelines on the duraark-system Github pages, starting with installing git, Docker, docker-compose and doing a git pull to get the initial download of content in order. After that we used the command "docker compose up -d" to start the main downloading of DURAARK system. We have a good internet connection and it took about an hour or so to download this. Then it was a matter of starting Firefox and connecting to localhost and from there on it worked nicely. The only snag was that we had a really old version of Firefox (version 32) that needed updating so that it was compatible with modern web technologies. With a newer one (version 42) it worked well."

¹¹Docker Setup (Ubuntu Linux): <http://docs.docker.com/engine/installation/ubuntu/>

5.2 Workbench UI - Functionality

Aspects evaluated

This action assesses the functionality of the WorkbenchUI based on a qualitative evaluation of a questionnaire taken by three stakeholders. The evaluated aspects are the applicability of the GUI for the user requirements defined in D2.2.1 and the user experience with the design of the graphical user interface.

Method

WP2 created a questionnaire¹² which contains questions regarding the workflow and the graphical user interface itself, detailed questions on each workflow step and on the potential to integrate WorkbenchUI functionality into the systems the stakeholders are working with. For each workflow step questions were asked on how helpful the current functionality is, how the user experience is and what should be enhanced.

This evaluation was done with a minimal set of stakeholders and it should be seen as a qualitatively interpreted experience report from people using the WorkbenchUI. We did not do a representative quantitative evaluation of the WorkbenchUI due to the prototype nature of the graphical user interface. Suggestions and comments made by the sample of stakeholders were used in v1.0.0 of the WorkbenchUI to enhance the user experience.

Datasets and Evaluation Activities

The questionnaire was taken by three expert stakeholders at CITA's premises. CITA was contextualizing the DURAARK project and the WorkbenchUI before the stakeholder tested the different parts of the WorkbenchUI. For this purpose CITA developed a slideshow, which introduced the stakeholders to the research approach of DURAARK and the workflows in which the WorkbenchUI would be embedded, see fig. 3. The whole document can be found in the appendix of this Deliverable¹.

After this the experts went through the WorkbenchUI and answered the questions. The three expert stakeholders were:

¹²WorkbenchUI Questionnaire: https://docs.google.com/spreadsheets/d/1MC_NE8oC0-4birub4tY_ZFYaoCmN7L-fiE-RNy4tuX4/edit#gid=1874743130&vpid=A1

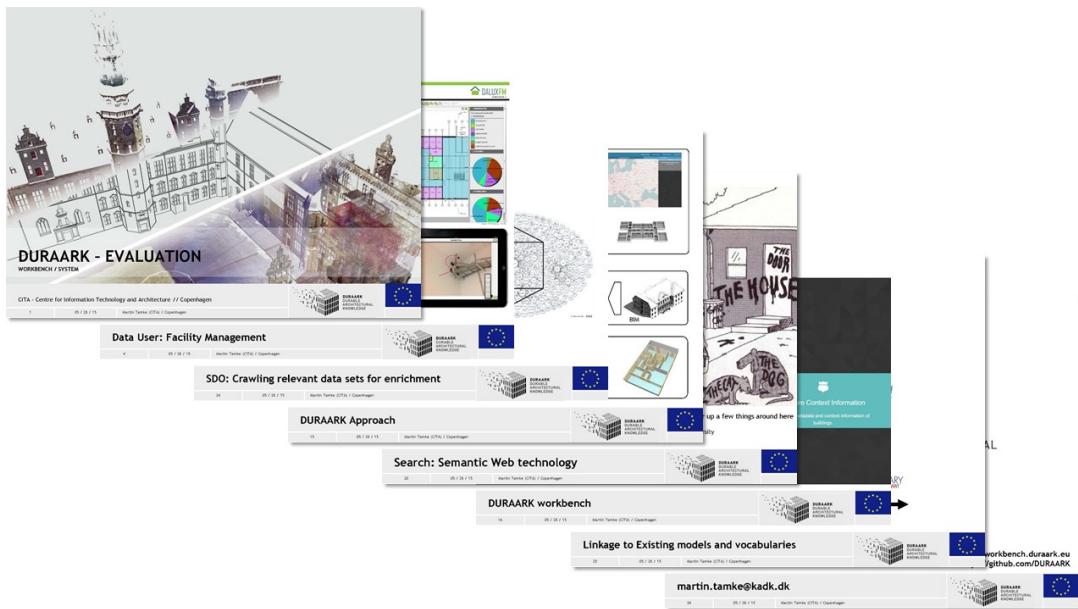


Figure 3: The WorkbenchUI questionnaire was contextualised with the help of an introductory slideshow.

Markus Lampe¹³ BIM Manager at Technical University of Denmark (Facility Management)

Esben Claussen Nørgard¹⁴ Research Assistant - Architect at KADK

Per-Kristian Hansson Architect, Malmö Sweden

Results

The raw result of the questionnaire is available at https://docs.google.com/spreadsheets/d/1MC_NE8oC0-4birub4tY_ZFYaoCmN7L-fiE-RNy4tuX4/edit#gid=1874743130&vpid=A1.

The next paragraphs assess the results of the main sections of the questionnaire.

The overall result of this evaluation is that the functionality provided in DURAARK is well received by the stakeholders and is seen as useful to them. All stakeholders would tightly integrate the functionality into their own work environment. The WorkbenchUI as graphical user interface has to be improved. The results of the evaluation showed that DURAARK is providing functionality stakeholders need and want to use, but the user interface is not yet transporting the full capabilities of the system through to the stakeholder. This is due to the prototype nature of the WorkbenchUI and gives a good

starting point for post-project development activities, i.e. the "Durable Building Data Association (DBD)" (see D8.7). The DBD will pick up the most pressing user interface issues and update the WorkbenchUI accordingly. Some of the user interface issues were already addressed and are available in v1.0.0 of the WorkbenchUI.

Workflow and Graphical User Interface This section evaluated the provided workflow and the applicability of the GUI for the use cases defined in D2.2.1. The expert stakeholders found the general workflow of archiving, searching and retrieving of data useful to work with. UC5 (Monitor the evolution of a structure over time), UC6 (Identify similar objects within a point-cloud scan to get an overview of interior objects) and UC9 (Enrich a BIM/IFC model with metadata from a repository) got the lowest applicability rates. This correlates with the results from the user interface sections which are responsible to carry out these use cases. For future use the user interface has to improve, not at least in order to make the functionality which DURAARK is providing better available. Since the time of the questionnaire and the writing of this document the user interface for UC5 (Monitor the evolution of a structure over time) and UC9 (Enrich a BIM/IFC model with metadata from a repository) have been improved in the WorkbenchUI.

All stakeholders are using network file storage for their 3D files, which is well covered by the system.

Ingest Buildings (Pre-Ingest Workflow)

Files The Files screen was perceived as sufficient for uploading files into the system. The user interface should be improved in providing a native file upload dialog additionally to the existing drag'n'drop functionality. One stakeholder mentioned that the possibility of adding 2D and PDF files to a session is lacking, which was not part of the project scope.

Metadata The amount of metadata fields should be larger. In the stakeholder's day to day work the usual amount of fields to enter for a building are larger. The user interface for entering metadata should reflect the importance of fields better in reordering the items. Also, a more common naming convention should be used for the fields.

Geometric Tools The "Reconstruct BIM Model" functionality was perceived as very useful. Its user interface has to be improved to transport the meaning of the functionality better. The "Detect Power Lines" functionality was very useful for the BIM manager, but not for the two architect stakeholders. The "Extract Floor Plan and Room Information" functionality was perceived as useful, however, the user interface was causing confusion in how the information was displayed. As a consequence we have merged the "Detect Power Lines" and the "Extract Floor Plan and Room Information" functionality into a single, more coherent user interface.

In general the functionality provided by the "Geometric Tools" is useful to stakeholders, but the user interface has to be improved.

Context Information The general idea of the "Context Information" functionality was well received by the BIM manager, but less useful for the architect stakeholders. Again, the user interface was causing confusion and did not transport the intended functionality well. We therefore redesigned the interface based on the suggestions in the evaluation to make it more useful.

Digital Preservation The "Digital Preservation" functionality is very useful to the stakeholders. The user interface is straightforward to use, but could show more information on the actual preservation content. This feature request was already integrated into the Workbench v1.0.0.

Search Buildings (Retrieval Workflow) This functionality was perceived as useful to the facility manager stakeholder, but less useful for the architects. From the user interface point of view a combination of the provided filters was a feature request, as well as a filtering based on production years. Another important feature request was to show different versions of the building. This feature request was implemented as part in adding the "Difference Detection" functionality in the "Geometric Tools" section of the WorkbenchUI in v1.0.0.

Explore Buildings and Advanced Search The "Explore Buildings" and the "Advanced Search" functionality were both in a proof-of-concept stage during the evaluation. In the WorkbenchUI v1.0.0 both functionalities have been directly integrated into the

“Search Buildings” page to simplify the user interface. Suggestions made in the questionnaire were taken into account.

5.3 Service Platform – Integratability

Integration into Stakeholder webplatform – BIMsync

Aspects evaluated Integrateability: Showcase ‘bimsync’

Figure 4 shows the integration layer used in this evaluation.

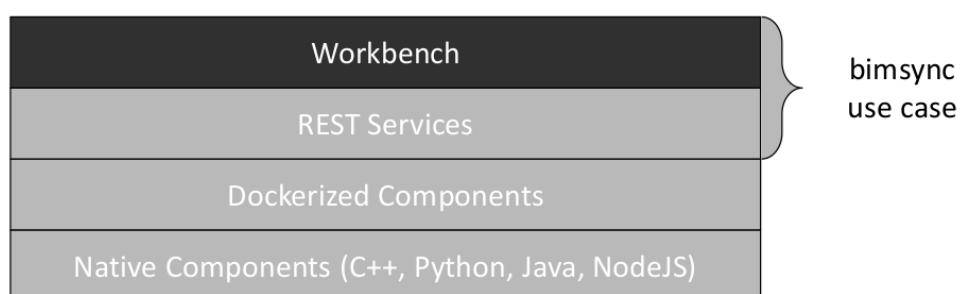


Figure 4: Catenda was evaluating the *REST Services* layer and the *WorkbenchUI* layer of the DURAARK System to integrate functionality into their BIMsync environment.

Method "1. Monitoring and evaluation of community interaction / community impact
2. Usability report from Catenda"

Integrating bimsync and DURAARK using the api from both

The DURAARK workbench has a powerful API to connect with external systems. The same is true for bimsync, the model server from DURAARK partner Catenda. Catenda created an integration with bimsync that is shown in the figure below.

The application was simple to create, and is available at <http://duraark.catenda.com>. It uses the session/submit API from Duraark and the viewer and data APIs in bimsync.

Datasets and Evaluation Activities Integrating bimsync¹⁵ functionality was easily done since there are clear and simple instructions available both for the bimsync part and

¹⁵<https://bimsync.com>

for the DURAARK part. How a REST client can use DURAARK is explained clearly on a wiki page on Github.¹⁶. The necessary steps are explained by example using the Curl¹⁷ tool, which is very popular on unix-like systems. It is also available on MS Windows.

The DURAARK system itself and how it is composed of microservices is explained on the wiki ¹⁸ part of the systems Github page.

Integrating bimsync into the same application was simple. The bimsync data and viewer API was used in the integration.¹⁹ To create a new application based on bimsync you need to contact Catenda and get access keys for API access. The bimsync API information explains step by step with REST examples how an application can ask a user for access, how to query a model for information, and how to do CRUD²⁰ operations. It also shows how to easily use the viewer API which enables high performance hardware accelerated 3D presentation / interaction with the BIM. Based on this information a nice new web application was developed (see below).

Results Figure 5 shows a new web application that uses the data- and viewer API from bimsync and the API from the DURAARK system. It shows an IFC-model that exists in bimsync, and is shown on top of terrain and property information streamed from an open data source (from the Norwegian mapping authority; Kartverket²¹).

A user accesses this application by its url, logs in using his username/password and selects one of his BIMs. From here on the user can click a button to initiate an archival process using the DURAARK system. In this example the user can manually enter the address of the building (see Figure 5), further user interaction would of course be simple to add. When the systems have done their work a message indicating success is shown to the user.

¹⁶<https://github.com/DURAARK/duraark-installer/wiki/Integration-via-the-DURAARK-Service-Platform>

¹⁷<http://curl.haxx.se/docs/manpage.html>

¹⁸<https://github.com/DURAARK/duraark-system>

¹⁹<https://bimsync.com/developers/reference/api/1.0>

²⁰https://en.wikipedia.org/wiki/Create,_read,_update_and_delete

²¹<http://kartverket.no/en/>



Figure 5: Catenda integrated the bimsync API and the DURAARK API.

Datasets and Evaluation Activities Integrating bimsync²² functionality was easily done since there are clear and simple instructions available both for the bimsync part and for the DURAARK part. How a REST client can use DURAARK is explained clearly on a wiki page on Github.²³. The necessary steps are explained by example using the Curl²⁴ tool, which is very popular on unix-like systems. It is also available on MS Windows.

The DURAARK system itself and how it is composed of microservices is explained on the wiki²⁵ part of the systems Github page.

Integrating bimsync into the same application was simple. The bimsync data and viewer API was used in the integration.²⁶ To create a new application based on bimsync you need to contact Catenda and get access keys for API access. The bimsync API information

²²<https://bimsync.com>

²³<https://github.com/DURAARK/duraark-installer/wiki/Integration-via-the-DURAARK-Service-Platform>

²⁴<http://curl.haxx.se/docs/manpage.html>

²⁵<https://github.com/DURAARK/duraark-system>

²⁶<https://bimsync.com/developers/reference/api/1.0>

explains step by step with REST examples how an application can ask a user for access, how to query a model for information, and how to do CRUD²⁷ operations. It also shows how to easily use the viewer API which enables high performance hardware accelerated 3D presentation / interaction with the BIM. Based on this information a nice new web application was developed (see below).

Figure 5 shows a new web application that uses the data- and viewer API from bimsync and the API from the DURAARK system. It shows an IFC-model that exists in bimsync, and is shown on top of terrain and property information streamed from an open data source (from the Norwegian mapping authority; Kartverket²⁸).

A user accesses this application by its url, logs in using his username/password and selects one of his BIMs. From here on the user can click a button to initiate an archival process using the DURAARK system. In this example the user can manually enter the address of the building (see Figure 5), further user interaction would of course be simple to add. When the systems have done their work a message indicating success is shown to the user.

Results Evaluating the effort of integrating DURAARK and bimsync shows that it was simple and required a relatively short amount of time. It was carried out in a time spanning a few days. There were no main obstacle as both APIs are simple to use and they work well together. The level of knowledge required is that the person who want to integrate either (or both) of these systems with their own system has a basic knowledge of web development. The keys are knowing about REST-based web services, javascript and HTML, and then following examples in the documentation.

A nice way to start the implementation is to get the main steps working using a tool like curl. Then it is a question of putting this demonstrated functionality into your favourite or available programming language on the server and serve the resulting webpage to the user.

²⁷https://en.wikipedia.org/wiki/Create,_read,_update_and_delete

²⁸<http://kartverket.no/en/>

Integration into Stakeholder work environment – Rhino3D

Aspects evaluated The evaluation targets stakeholders from the group of Data Users and investigates the efforts to implement the DURAARK system and its components into a work environment of this group.

Figure 6 shows the integration layers used in this evaluation.

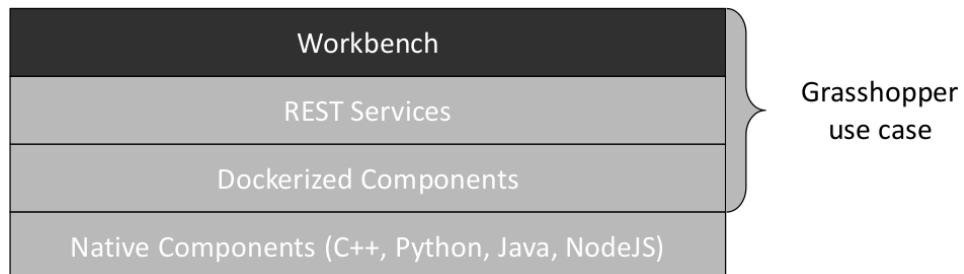


Figure 6: CITA was evaluating the *REST Services* layer and the *WorkbenchUI* layer of the DURAARK System to integrate functionality into their Rhino3D environment.

Method Testing the integratability of the DURAARK platform in the working environment of the stakeholders was a main focus of this evaluation activity. Through the use of existing and the development of a new Point Cloud plugin CITA created a prototypical implementation of the DURAARK platform in Rhino 3D, see figure 7. The Rhino CAD environment was chosen as a platform for software development, thanks to it's users' activity (over 40,000 registered users) and variety of their specializations, online documentation and extensibility. Moreover, Rhino is a fully fledged BIM Modeller through the VisualArq add-on. Point Cloud functionality was added through the Volvox plug-in, developed by CITA. Volvox utilises the parametric modelling add-on for Rhino – Grasshopper – which was chosen as the technical basis for Volvox. The openness and extendability of Grasshopper (over 140 various add-ons), enabled Volvox to utilize the Rhino 3D/Grasshopper environment in a very concentrated and optimized way. Volvox implements both retrieval and ingest part of the workbench, enabling the users to create, edit and update the workbench data. It also implements Docker for IFC reconstruction, and cooperates with VisualArq for BIM manipulation.

Datasets and Evaluation Activities As Grasshopper is a node-based visual programming IDE, the Volvox plug-in consists of multiple components (nodes). To establish

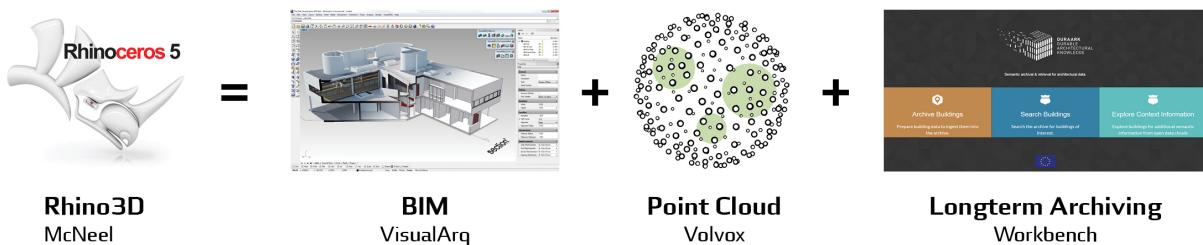


Figure 7: Stakeholder work environment for integration efforts

connection with the workbench, user has to use one of the components, which comes with the predefined required inputs. Once the connection is established, it's possible to download and edit e57 point clouds (Volvox) or IFC files (VisualArq) and upload it after editing with one of the Volvox components.

Results It took one month to develop and release the Volvox (version 0.1.0.0), with a basic point cloud manipulation toolset. After 4 more months of development, a grand update was released (version 0.2.0.0), which brought multiple changes and new tools to Volvox: multi-threading, code optimization, workbench and IFC integration. One of the biggest problems encountered during the development was lack of .NET libraries for e57 file format reading and writing. The widely used library (libE57) is written in C++ which is not easy to integrate with Grasshopper (which itself is written in .NET as well). A new library was developed – E57Lib, which is the first .NET E57 read/write library. The library is now published on the DURAARK GitHub.

Another problem stems from the very nature of Grasshopper: almost every node creates a deep copy of its input data, which is particularly undesired when operating with big point clouds (lack of RAM memory). It was overcome by creating a procedure-based node which can combine multiple nodes internally, based on the set of input instructions – that means that the data is copied only once.

The Duraark Workbench Docker container was relatively simply to integrate. Initial problems could be discussed with WP2 and led to the development of a Duraark Windows Installer for the Docker container. This made the installation easy and user-friendly. A few corrections in the individual component batch files make sure, that the deployed program receives a structured and simplified workflow with writing and reading files from a predefined folder structure. The .NET code for the Grasshopper components could be

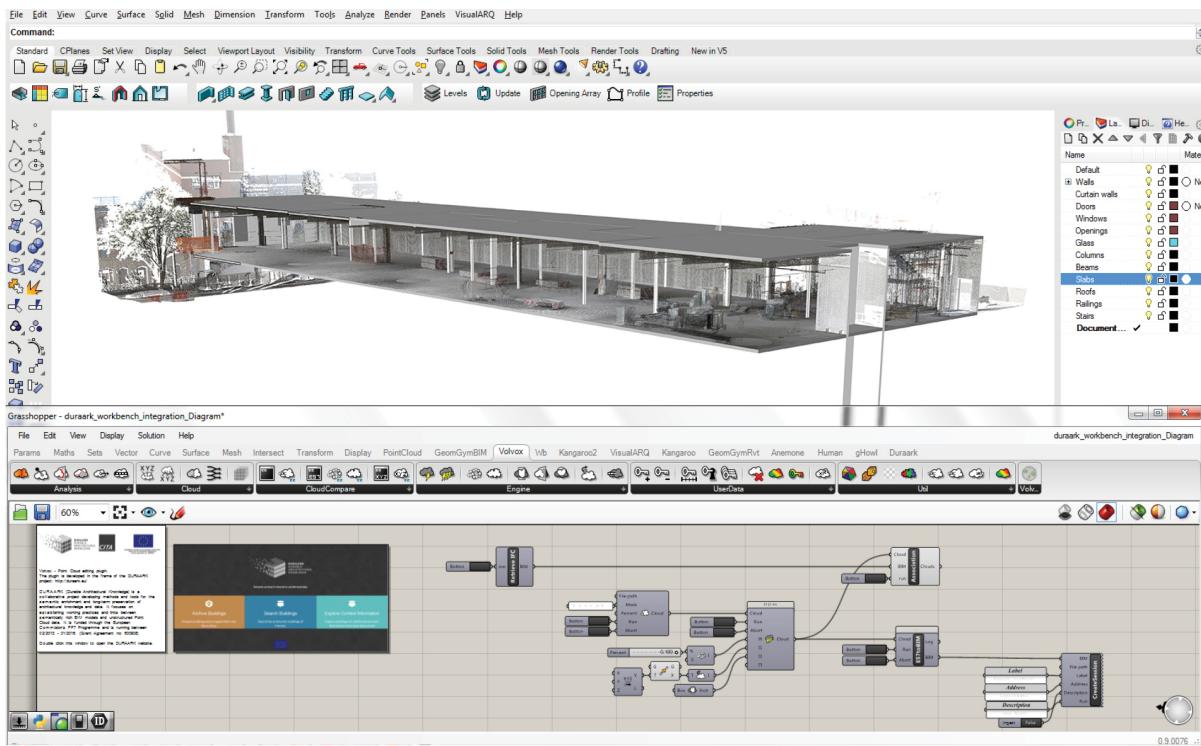


Figure 8: Rhino3D and Grasshopper with VisualArq, Volvox and the Workbench.

written with relative little effort. The Grasshopper components run the Docker Containers and simultaneously receive and display standard output and error messages.

The efforts for this implementation was approximately one month, counting in the iterations with WP2, which in turn led to the duraark-installer for Windows.

Integration into Library environment - Metadata extractors (IFC, E57)

Aspects evaluated TIB is hosting the Rosetta DPS on a Solaris operating system. One future goal of TIB is to directly integrate the metadata extractors implemented in DURAARK into the Rosetta system (currently the metdata extraction is done in the DURAARK System and the results are ingested into Rosetta as part of a SIP). To integrate the metadata extractor components *pyIfcExtract* for IFC and *e57-metadata* for E57 files the genuine way would be to include them via their native implementation as Rosetta plugins. However, Solaris is not a supported operating system by the DURAARK System, due to its rare spreading²⁹. Therefore it is not possible to integrate the extractors

²⁹ Additionally, TIB will change the operating systems to a Linux based solution in the future.

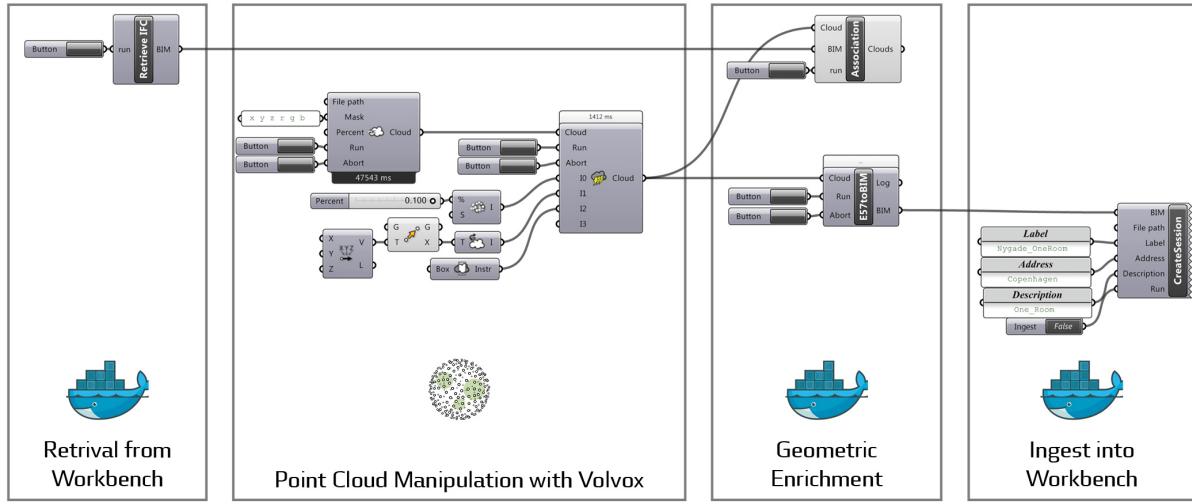


Figure 9: This diagram shows a section of the DURAARK Rhino/Grasshopper integration components. It shows the steps from retrieving an iFC model from the Workbench through manipulating the data with Volvox and the geometric enrichment components to re-ingesting the data to the Workbench..

either via Docker or via their native implementation.

This activity provides a proof-of-concept evaluation on using the the "duraark-metadata" REST service layer to integrate the metadata extraction parts of DURAARK's functionality into the Rosetta system, due to the fact that the dockerized and the native component integration is no option here. Goal of this evaluation activity is to prove that the developed metadata extractors could be used by Rosetta as an external service (with respect to Rosetta's system boundaries).

Figure 10 shows the integration layer used in this evaluation.

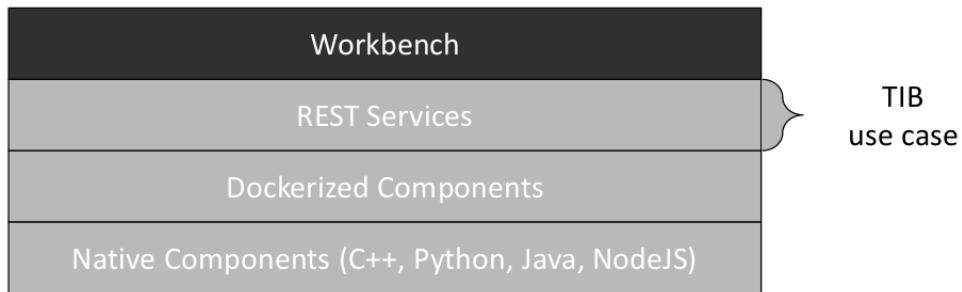


Figure 10: TIB was evaluating the *REST Services* layer of the DURAARK System to integrate functionality into their Rosetta DPS in a proof-of-concept evaluation.

Method Selected datasets will be ingested into Rosetta with the technical metadata (ifcM and e57M) extracted by the "duraark-metadata" REST service. The data is then ingested into Rosetta and the result validated against the known ground truth of the technical metadata contained in the datasets.

Datasets and Evaluation Activities The datasets and their evaluation is described in Section 8. There the SIP structure of test datasets is evaluated, where the metadata contained in the SIPs was generated by the "duraark-metadata" REST service. Therefore the evaluation in Section 8 directly applies here, too.

Results Using the "duraark-metadata" service allows the extraction functionality to be integrated into the Rosetta system via an external web service approach. Rosetta requests metadata for files via the "duraark-metadata" REST service, which is running as an independent system on TIB's premises. The underlying connection between the system is the file storage layer. Rosetta schedules the extraction of a certain file path at the "duraark-metadata" service, the service is accessing the file path, extracts the metadata and returns it to the calling Rosetta system.

The "duraark-metadata" API is stable and provides the functionality needed by Rosetta to get access to the file's metadata. Updates to the "duraark-metadata" service can be easily tested before going into production. It is also not necessary to restart the Rosetta system for updates of the metadata extraction service.

6 Evaluation - Semantic Enrichment Components

This part of the Evaluation Deliverable investigates the tools and approaches, that are concerned with the semantic enrichment of building data. This includes software components, that are part of the pipeline for the DURAARK WorkbenchUI, such as the **Focused Crawler** (see section 6.1) or the **IFC metadata extractor**(see section 6.2), components for the DURAARK system such as the **Entity Retrieval** (see section 6.3) and the quality that developed tools achieve for the automated matching and linking of **Vocabularies**. Furthermore, the evaluation of the **IFC Point Cloud Schema** (see 6.4) and the integrated **IFC HDF5 Serialization** (see 6.5) linking Point Clouds and BIM files which is in the process of standardization, as described in D8.7.

6.1 Enrichment (Focused Crawler)

In D3.6, we introduce a focused crawler for linked data, which replaces the previously developed crawling environment with a more targeted and hence scalable approach. Based on experimentally defined crawl configurations, we introduce an efficient means to crawl linked data of relevance to the specific instances in the SDA. In this Section we describe the evaluation of the focused crawler, which is a means for semantic enrichment of archival data. In addition we evaluate the entity retrieval framework within the SDA.

Aspects Evaluated

We evaluate the precision of the focused crawling approach. We first generate a ground truth using crowdsourcing (where crowd workers assess the relevance of crawled data to a given instance). Next, we analyze the performance of the proposed focused crawling framework. This addressing the stakeholders interested in usecases stemming from the semantic enrichment of architectural data to be archived.

Experimental Setup: Method and Datasets

In the experimental setup, we introduce the datasets we consider for the crawling process and the approaches on generating the seed lists. Next, we assess the performance of the different crawling configurations, constrained for the varying parameters: (i) *seed list coherence*, (ii) *hop-size* and (iii) *crawl-candidate ranking approaches*. Consequently, we

draw conclusions about the optimal *crawl parameters*. Finally, we compare our approach against existing baselines.

Seed Lists Given that there are no existing datasets with *seed lists* for the focused crawling process, and since it simply represents an ad-hoc information need for a specific topic, we motivate the seed list generation based on the document annotation scenario, where seed entities are directly linked to and extracted from documents in a corpus. In order to eliminate noise, potentially introduced by extracting entities from an arbitrary corpus, we directly utilise Wikipedia, where DBpedia entities can be derived directly from the hyperlinks of a source page to other Wikipedia pages, i.e. the equivalent of the Wikipedia entity http://en.wikipedia.org/wiki/Cologne_Cathedral is http://dbpedia.org/resource/Cologne_Cathedral.

We generate a sample of seed lists that are used to assess the crawling process for the parameters, such as the *seed list coherence*, *crawl intent* etc. The heuristic we rely on to generate the seed lists is to use the Wikipedia page view statistics³⁰. The page views are gathered by real-world users that visit the different entity pages. In this way, our starting point for generating seed lists is the information need by the Wikipedia users, namely the top visited entity pages.

Since popular Wikipedia pages tend to be linked more consistently, we consider the top entities based on the page views from the first week of 2012, which consists of 1.6 billion page views. We have specifically chosen a period in the past to ensure well-populated and popular pages, rather than new and insufficiently annotated pages.

Next, to generate seed lists of varying *coherence*, we follow two different strategies.

1. **High-Coherence:** top- k entities from a single entity page
2. **Low-Coherence:** top- k entities extracted from entity pages that are related to the top-viewed entity page based on the Wikipedia categories

The intuition is that (1) will produce more coherent and (2) more diverse seed lists. In (1) we start with a top-viewed entity page u and extract the entities $\{u_1, u_2, \dots, u_t\}$ that appear in the entity page (entities are marked by *anchors*) and rank according to their frequency. Next, we take the top- k frequent entities appearing in u , assuming that such entities are more closely related to u . We use the top- k entities as seeds.

³⁰<https://dumps.wikimedia.org/other/pagecounts-raw/>

For (2) we start again with a top viewed entity page in Wikipedia. For each entity u , we randomly obtain the category from the set of categories associated with u and retrieve all entities $\{u_1, u_2, \dots, u_k\}$ in the same category. We then extract the entities $\{u'_1, u'_2, \dots, u'_k\}$ appearing in the pages of $\{u_1, u_2, \dots, u_k\}$ as candidates. Finally, a random sample of k entities from the accumulated candidate set are considered as low coherence seed list. For this process, we use top-3 Wikipedia entity pages and three Wikipedia categories. As a result we obtain six seed lists with varying coherence Γ . The average coherence scores for the *high coherence* seed lists is $\Gamma = 0.55$, whereas for the *low coherence* seeds we have $\Gamma = 0.34$. The seed lists and the corresponding ground-truth are available for download³¹.

Crowdsourced Ground Truth We leverage crowdsourcing as a means to establish the ground truth in the form of a ranking of the most relevant entities within a specified n -hop neighborhood for a given seed list S . We adopt the following steps in order to establish the ground truth for evaluation.

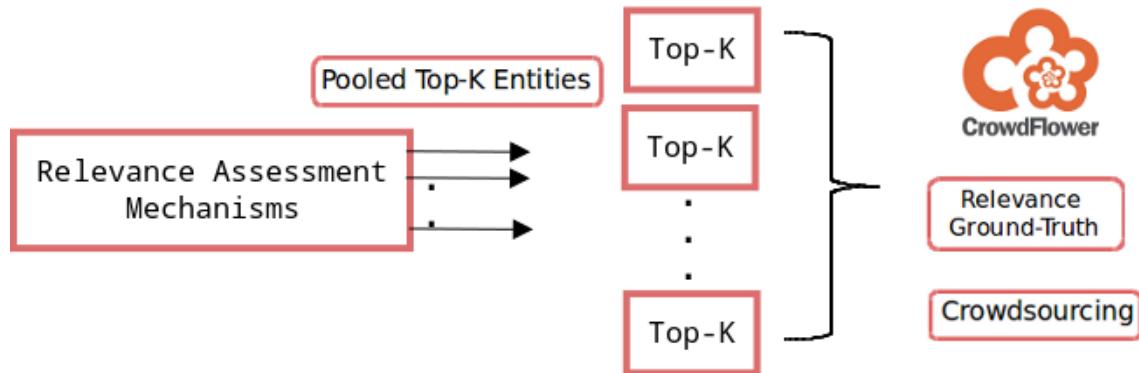


Figure 11: Generation of ground truth for evaluation of relevance assessment measures.

First, we pool the top- k entities³², as obtained by using the different relevance assessment measures. We model the activity of assessing the relevance of an entity to the seed list as an atomic microtask that can be deployed on a crowdsourcing platform such as CrowdFlower. In order to assist the crowd workers in assessing the relevance of an entity to a given seed list, we represent the seed list with a corresponding *topic* that describes the collective meaning projected by the seed list. In order to determine an apt *topic* that can represent a seed list and thereby the crawl intent, we follow a gamified approach

³¹http://13s.de/~fetahu/crawler_wise2015/

³²In this case we pooled the $Top - 500$ entities resulting from all the different configurations for each seed list.

such as the popular ESP game proposed by Von Ahn and Dabbish [?]. Each seed list is presented to five different experts, who are then asked to independently tag the seed lists with representative topics. When all five experts reach a consensus with respect to a particular description, that tag is chosen to be the representative *topic* of the seed list. For instance, for the seed list generated from How I Met Your Mother (TV Series) Wikipedia page, experts assigned the topic *How I Met Your Mother (TV Series) - Characters and Actors*.

Next, we present each entity gathered from the neighborhood of the seed list alongside the corresponding *topic* to crowd workers and request judgments of relevance according to a 5-point Likert-scale (ranging from *Not Relevant* to *Highly Relevant*). By aggregating the judgments from the workers into relevance scores, and ranking the pooled top- k entities based on these relevance scores, we thereby generate our ground truth of relevant entities corresponding to the seed list S . Finally, we use the ground truth thus established to evaluate each of the relevance assessment methods across the top- k entities.

In order to ensure that the ground truth is reliable, we take several precautions as recommended by our previous works [6, 5] to curtail malicious activity and to avoid misinterpretations in the relevance scoring.

Crawl Configurations and Baselines A *crawling configuration* is defined by (a) candidate ranking approach, (b) hop size (depth) for the crawling process, and (c) seed list coherence. Here the candidate ranking considers two cases: (i) candidate relevance scoring taking into account the attrition factor λ which is denoted as $EJSim^+$, and (ii) candidate relevance scoring without the attrition factor, i.e. the baselines below. The maximum hop size *depth* considered during the crawling process is one of the most significant impact factors for runtime and NDCG score. We run experiments with *depth* in {2, 3}, the corresponding methods are denoted with a subscript indicating the depth, where, for instance, $EJSim_2$ indicates a *depth* = 2.

We consider the following baseline and state-of-the-art approaches for the focused crawling task.

- **EJSim.** Entity Jaccard Similarity, $EJSim$ computed as $EJSim^+$ but without considering the *attrition factor*.

- **NJSim.** We also consider the graph-based relatedness baseline, the *Neighborhood Jaccard Similarity* which is computed as follows.

$$\text{PairwiseNJSim}(e_i, c_j) = \frac{1}{\lambda(e_i)} \cdot \sum_{k=1}^{\tau} \alpha^k \frac{|N_k(e_i) \cap N_k(c_j)|}{|N_k(e_i) \cup N_k(c_j)|} \quad (1)$$

$$\text{NJSim}(c_j) = \frac{\sum_{i=1}^n \text{PairwiseNJSim}(e_i, c_j)}{n} \quad (2)$$

where, $N_k(e_i)$, $N_k(c_j)$ are the sets of neighboring vertices at the k^{th} step of seed entity (e_i) and candidate entity (c_j) respectively, $\alpha \in [0, 1]$ is a real number to ensure that closer neighbors have lower weight, and τ is the maximum length of the paths considered.

- **PageRank.** Is a widely adopted approach for ranking crawling results (documents or entities). The computation of PageRank is formalized in Equation 3.

$$\text{PageRank}(c_i) = \frac{1-d}{N} + d \sum_{c_j \in M(c_i)} \frac{\text{PR}(c_j)}{L(c_j)} \quad (3)$$

where $M(c_i)$ is the set of entities linked to candidate c_i in the entity graph, $L(c_j)$ is the number of outbound links to the entity c_j , N is the total number of entities, and d is the damping factor [1].

- **NB.** We consider [2] as state-of-the-art approach for focused web crawling, which we adopt for our experimental setup. However, while this approach originally implements focused crawling of Web documents as opposed to Linked Data, we introduce some adaptations. The DOM tree features cannot be considered in case of Linked Data, hence, we adopt lexical features for the NB classification. The state-of-the-art approach uses predefined topic taxonomy with example URLs as training set. In our scenario, the topic, which is called crawl intent in our work, is dependent on the seed list without predefined limitations. However, our groundtruth can be used as training set for a classifier. The classification of candidate entity as relevant is performed by the function formalized in Equation 4.

$$\text{NB}(c_i) = P(y = 1|c_i) = \frac{P(y)P(c_i|y)}{P(c_i)} \quad (4)$$

where $P(y = 1|c_i)$ is the probability of candidate c_i belongs to class $y = 1$, which in our case translates as a ‘relevant’ entity for a given seed list.

Evaluation Metrics To evaluate the crawling performance of the different configurations, we consider the *Normalized Discounted Cumulative Gain (NDCG)* of the ranked set of candidate entities with respect to an established ground truth. The NDCG score is computed as follows:

$$nDCG@k = \frac{DCG@k}{iDCG@k} \quad DCG@k = rel_1 + \sum_{i=2}^k \frac{rel_i}{\log_2 i}$$

where, $DCG@k$ represents the discounted cumulative gain at rank ‘ k ’, and $iDCG@k$ is the ideal $DCG@k$ computed with respect to the *ground truth* (described in the following section).

Another important factor to evaluate is the runtime performance of the different configurations. Due to the likelihood of large set of candidate entities existing for a particular seed list, runtime is a crucial factor. The runtime is simply measured in terms of the amount of time taken to complete a full-cycle of the crawling process.

Results

Performance of Focused Crawling Configurations In this section we report and discuss the performance of the different focused crawling configurations. On average, for the different seed lists we crawl approximately 491,425 triples. Hence, an important aspect is how well the different crawl configurations rank the candidate entities for their relevance w.r.t the seed lists. In Table 1 we report the performance of the different crawling configurations. The values are reported for the average NDCG scores at rank 100.

Coherence	$EJSim_2^+$	$EJSim_2$	$NJSim_2$	NB_2	$PageRank_2$
$\Gamma < 0.5$	0.7446	0.7364	0.5752	0.6955	0.5346
$\Gamma \geq 0.5$	0.6876	0.6802	0.5193	0.5273	0.4408
	$EJSim_3^+$	$EJSim_3$	$NJSim_3$	NB_3	$PageRank_3$
$\Gamma < 0.5$	0.6928	0.6382	0.6608	0.6612	0.5722
$\Gamma \geq 0.5$	0.5831	0.5740	0.5197	0.4821	0.4166

Table 1: Average NDCG@100 for different focused crawling configurations across seed lists with varying coherence.

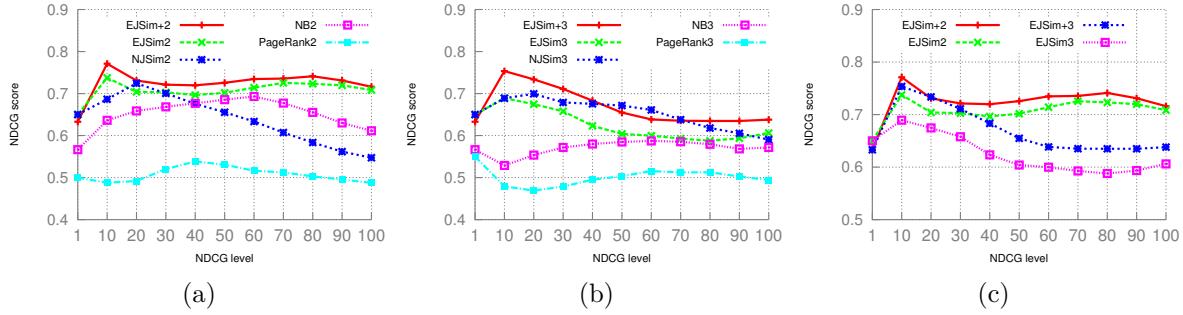


Figure 12: (a) Performance of different configurations with depth 2 measured as average NDCG across all seed lists. (b) Performance of different configurations with depth 3 measured as average NDCG across all seed lists. (c) Performance of different configurations (depth 2 and 3) measured as average NDCG across all seed lists.

- **Performance.** Table 1 presents the average NDCG@100 score of different configurations described in Section 6.1 with seed lists of varying coherence. Figure 12a and 12b present the NDCG score at different levels for the different configurations. Figure 12c presents a detail comparison between the performance of *EJSim* and *EJSim*⁺. It shows that the NDCG scores increased after introducing the attrition factor to *EJSim*. The average improvement across different NDCG levels is 1.6% on *depth 2* and 4.3% on *depth 3*. This indicates that the *attrition factor* has a positive effect. The *coherence* of the seed list in our experimental setup does not have a significant impact on the focused crawling configuration setups.

From the results we can also observe that the *EJSim*⁺ outperforms baseline methods, specifically *NJSim* and *PageRank*. The NDCG@100 of *EJSim*⁺ is 16.9% and 4.8% higher than *NJSim* for *depth 2* and *3* respectively. The NDCG scores at different levels (Figure 12a and 12b) also support this insight. *PageRank* shows the weakest performance since it does not take crawl intent into consideration during candidate ranking. Furthermore, *EJSim*⁺ has 7.3% and 9.8% improvement in average for *depth 2* and *3* respectively compare to *NB* in our experiment. Based on the dataset, one of the reasons that *NB* and *EJSim*⁺ have different performance while both use lexical features is that the classes within the training set are not independent from each other. The overlap between the feature set of different classes causes negative effects on the *NB* classification. The reason for this overlap is that the crawl intent is based on the seed lists instead of the topic from a predefined

topic taxonomy. This also reveals that our method is more robust and flexible for variety of crawl intent.

Another useful and conclusive insight is the fact that *depth* 3 has not led to performance gains with respect to *crawl depth* of 2. In fact we can observe the opposite from Figure 12c, where for *EJSim⁺*, increasing the crawl depth beyond 2 seems to lead to weaker NDCG scores on average.

- **Efficiency.** Since the time-intensive step is the candidate crawling, there is no considerable difference between the crawling configurations with different ranking methods. However, the average runtime across seed lists and ranking methods increased from 548 seconds of depth 2 to 1936.6 seconds of depth 3 in average while there is no significant improvement of the NDCG score. Given the significantly increased runtime when crawling beyond hop 2, a crawl depth of 2 seems to provide optimal performance and is not advisable to crawl to a higher distance.

6.2 IFC metadata extractor

The IFC metadata extractor is a core part of the ingestion pipeline of the DURAARK WorkbenchUI. Out of an input IFC file, it is able to extract literal values and aggregates thereof to form metadata records that can be further curated by end-users. The output are instances of entities as defined in the buildm schema, which includes valid descriptive metadata for ifc and e57 files.

Aspects evaluated

The aspects that have been evaluated on account of this Deliverable are those that guarantee the correct functioning of the module within the DURAARK WorkbenchUI.

1. The **scalability** of the extractor is assessed in order to guarantee the system does not halt on large or complex files.
2. The **usefulness in preservation tasks** is a functional requirement to assess to what extent the tool and extracted data can contribute to common preservation tasks, such as preservation planning and quality assessment of the ingested data.

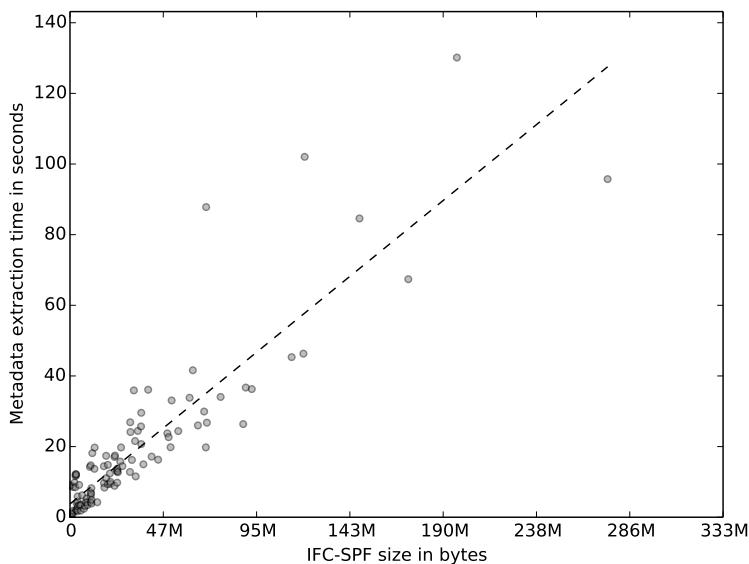


Figure 13: Time to process IFC-SPF files based on their file size

Method

The component is evaluated by feeding IFC files from the DURAARK dataset into the extractor, capturing the output to verify its integrity and taking note of performance characteristics of the system. Statistical performance measures are deduced in order to extrapolate the performance characteristics of the system.

Datasets and Evaluation Activities

The IFC files from the DURAARK dataset are used. This includes a total of 115 IFC-SPF files. Some of which are small variations of the same model, or describe different aspects of the same overarching structure.

Results

Scalability The execution time necessary for the extractor to finish processing a single file is recorded for all files in the dataset. By means of the statistical measures one can observe the extraction follows a linear time-complexity curve. This is also the theoretical optimum, as the entire file needs to be parsed. The slope of the line through the data indicates the extractor is able to consume 2.2 megabytes of data per second

on average³³, including the time to parse the IFC-SPF data and apply the extraction rules. This is considered competitive, especially considering evaluation of the extraction rules happens in an interpreted language (Python). Results are graphically depicted in Figure 13. Deviations from the fitted line are due to the exact population of entity instances. For example, the time to process geometrical properties such as `duraark:totalWindowArea` and `duraark:averageWallThickness` depends on the amount of instances of relevant entities in the model.

Usefulness in preservation tasks The applicability of the extractor and its output is judged in two distinct preservation tasks. Firstly, as part of a preservation planning activity the migration from IFC2x3 to IFC4 data is considered. More details about this scenario can be found in DURAARK Deliverable D6.3. In this task extracted IFC metadata is of great help to quickly determine the integrity and authenticity of migrated data. Excerpts from extracted data are listed in Listings 1 2 3. In one of the migration strategies (3), which entails reimporting the IFC2x3 data into Autodesk Revit in order to export IFC4, one can observe unacceptable change in the total window area, which demonstrates the infeasibility of this migration approach over the alternative of re-exporting IFC data from native files (2). Note that these excerpts feature only a small subset of the total amount of automatically extracted predicates.

Secondly, the sustainability of the extracted data as a means to assess the overall quality in a repository is assessed by means of the technical metadata that is collected as part of the ifcm schema. The ratio of `duraark:relationshipCount` and `duraark:instanceCount` is presented in Figure 14 and represents the amount of relational metadata contained in the model. Various semantics, such as association of materials, type traits, property characteristics and topological connectivity are all examples of the use of objectified relationships in the IFC schema, which is collected as a measure of semantic richness in the ifcm schema. The assessed models and their trend line show that on average the amount of relational constructs is higher for more recent models, but in addition, the variance is considerable.

³³This test is executed on a laptop with Intel Core i5-4200U processor and 4GB RAM, with a 5.400rpm SATA-600 hard drive.

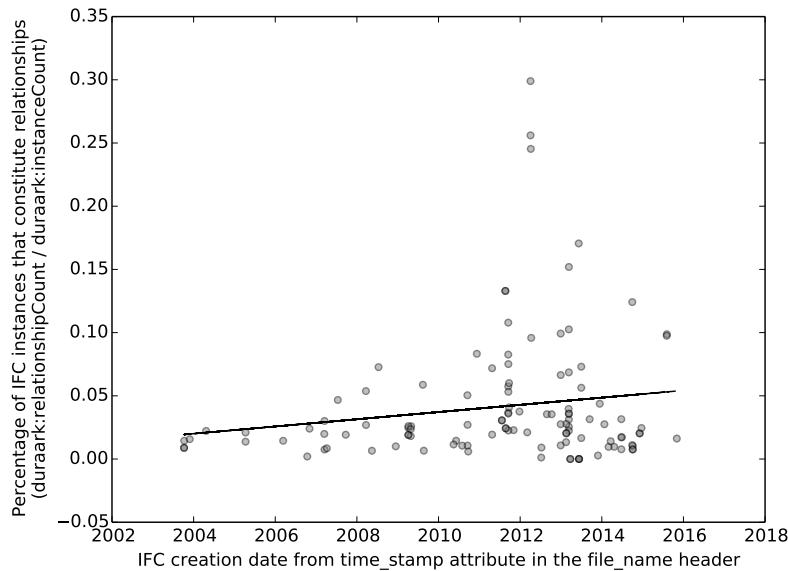


Figure 14: Semantic richness of IFC models over the years, assessed by the amount of relational metadata.

Listing 1: Excerpt of extracted data for an IFC2x3 model

```

<http://data.duraark.eu/resource/
  ↳ ifcspffile_660dd70ff34b41febde7599e23857e29> a
  ↳ duraark:IFCSPFFile ;
duraark:entityCount "96"^^xsd:nonNegativeInteger ;
duraark:filename "Projektnummer"^^xsd:string ;
duraark:hasType "Model"^^xsd:string ;
duraark:instanceCount "346311"^^xsd:nonNegativeInteger ;
duraark:optionalAttributesSet "0.527789982425"^^xsd:double ;
duraark:relationshipCount "2540"^^xsd:nonNegativeInteger ;
duraark:represents <http://data.duraark.eu/resource/
  ↳ physicalasset_660dd70ff34b41febde7599e23857e29> .

```



```

<http://data.duraark.eu/resource/
  ↳ physicalasset_660dd70ff34b41febde7599e23857e29> a
  ↳ duraark:PhysicalAsset ;
duraark:averageWallThickness "0.359255"^^xsd:decimal ;
duraark:totalWindowArea "241.211124"^^xsd:decimal .

```

Listing 2: Excerpt of extracted data for an IFC4 model

```

<http://data.duraark.eu/resource/
  ↳ ifcspffile_660dd70ff34b41febde7599e23857e29> a duraark:
  ↳ IFCSPFFile ;
duraark:entityCount "97"^^xsd:nonNegativeInteger ;

```

```

duraark:filename "Projektnummer"^^xsd:string ;
duraark:hasType "Model"^^xsd:string ;
duraark:instanceCount "346549"^^xsd:nonNegativeInteger ;
duraark:optionalAttributesSet "0.523205471141"^^xsd:double ;
duraark:relationshipCount "2616"^^xsd:nonNegativeInteger ;
duraark:represents <http://data.duraark.eu/resource/
    ↪ physicalasset_660dd70ff34b41febde7599e23857e29> .

<http://data.duraark.eu/resource/
    ↪ physicalasset_660dd70ff34b41febde7599e23857e29> a duraark:
    ↪ PhysicalAsset ;
duraark:averageWallThickness "0.359255"^^xsd:decimal ;
duraark:totalWindowArea "241.211124"^^xsd:decimal .

```

Listing 3: Excerpt of extracted data for an IFC2x3 model that is imported and exported as IFC4

```

<http://data.duraark.eu/resource/
    ↪ ifcspffile_660dd70ff34b41febde7599e23857e29> a duraark:
    ↪ IFCSPFFile ;
duraark:entityCount "93"^^xsd:nonNegativeInteger ;
duraark:filename "0001"^^xsd:string ;
duraark:hasType "Model"^^xsd:string ;
duraark:instanceCount "483268"^^xsd:nonNegativeInteger ;
duraark:optionalAttributesSet "0.53784860558"^^xsd:double ;
duraark:relationshipCount "2261"^^xsd:nonNegativeInteger ;
duraark:represents <http://data.duraark.eu/resource/
    ↪ physicalasset_660dd70ff34b41febde7599e23857e29> .

<http://data.duraark.eu/resource/
    ↪ physicalasset_660dd70ff34b41febde7599e23857e29> a duraark:
    ↪ PhysicalAsset ;
duraark:averageWallThickness "0.3321617"^^xsd:decimal ;
duraark:totalWindowArea "125.1885723"^^xsd:decimal .

```

6.3 Entity Retrieval (SDAS)

In this section we report evaluation results of the two main step of the entity retrieval (ER) framework as part of the SDA. The two main steps as part of the ER are: (1) entity clustering as a pre-processing step, and (2) entity retrieval evaluation. The retrieval framework is reported in detail in Deliverable D3.6 and in [4].

Aspects evaluated

We first evaluate the quality of the pre-processing step, i.e., the clustering results for the *x-means* and *spectral* clustering algorithms. Next, we present the findings from our rigorous evaluation on the entity retrieval task.

Method

For both aspects we rely on crowd-sourcing evaluation for the entity clustering and entity retrieval tasks. We follow well-established and rigorous guidelines. We ensure that the quality of the crowd-sourced results is high and reliable.

In order to ensure that we acquire reliable responses from the crowd workers, we take several precautions while designing the tasks for the evaluation of clusters, as well as establishing the ground-truth for the retrieval of entities. We provide clear instructions and examples to avoid misinterpretations in the relevance scoring, leading to a bias in the judgements. We compensate workers with monetary incentives that are proportionate to their contribution. In addition, we use *gold standard questions* as recommended by previous works to curtail malicious activity.

Datasets and Evaluation Activities

Dataset. We evaluate the entity retrieval framework on the BTC12 dataset [?]. BTC12 represents one of the largest periodic crawls of Linked Data, also containing well-known knowledge bases like Freebase and DBpedia. The overall statistics of the data are: (i) 1.4 billion triples, (ii) 107,967 graphs, (iii) 3,321 entity types, and (iv) 454 million entities.

Entity Clusters. The statistics for the generated clusters are as follows: the average number of entities fed into the *LSH* bucketing algorithm is 77,485, whereas the average number of entities fed into *x-means* and *spectral* is 400. The number of generated entity buckets by LSH is 20,2009, while the number of clusters for *x-means* and *spectral* is 13 and 38, with an average of 10 and 20 entities per cluster respectively.

Query Dataset. To evaluate our retrieval approach we use the *SemSearch*³⁴ query set from 2010 with 92 queries. The SemSearch query set is a standard collection for evaluating entity retrieval tasks.

Results

Cluster Accuracy Evaluation

Considering the large number of clusters that are produced in the pre-processing step for a given *type* and *bucket*, evaluating the accuracy and quality of all clusters is infeasible. We randomly select 10 entity types and 10 buckets, resulting in 100 clusters for evaluation, where for each cluster we randomly select a maximum of 10 entities.

To evaluate the *cluster accuracy*, we deploy atomic microtasks modeled such that a worker is presented with sets of 10 entities belonging to a cluster, along with a description of the entity in the form of the entity profile. The task of the worker is to pick the odd entities out (if any). We gather 5 judgments from different workers for each cluster. By enforcing restrictions available on the CrowdFlower platform, and following state of the art task design recommendations, we ensure that we receive judgments from the best workers (workers with high reputation as indicated by CrowdFlower).

Figure 15b presents our findings for the evaluation of the clustering process. We note that for *x-means* and *spectral* clustering approaches, nearly 35% and 38% of the clusters are judged to be perfect respectively (i.e., the entities within the cluster were all found to belong together). 39% of the clusters corresponding to *spectral* clustering and 40% of the clusters corresponding to *x-means*, have an accuracy of 80%. Considering its multidimensional representation of the entities, *spectral* clustering has higher accuracy and it does not have clusters below 70% accuracy. The lowest accuracy of 70% for *spectral* clustering implies that in each cluster there were only 3 entities that did not belong to the cluster. The implications of an accurate clustering process become clearer in the next section, where we assess the accuracy of finding relevant entities in the generated entity clusters.

Figure 15a presents the pairwise agreement between workers on the quality of each cluster. In case of the *spectral* clustering, we observe a high inter-worker agreement of 0.75 as per Krippendorf's Alpha. We observe a moderate inter-worker agreement of 0.6 as per Krippendorf's Alpha on the clusters resulting from *x-means*.

³⁴<http://km.aifb.kit.edu/ws/semssearch10/>

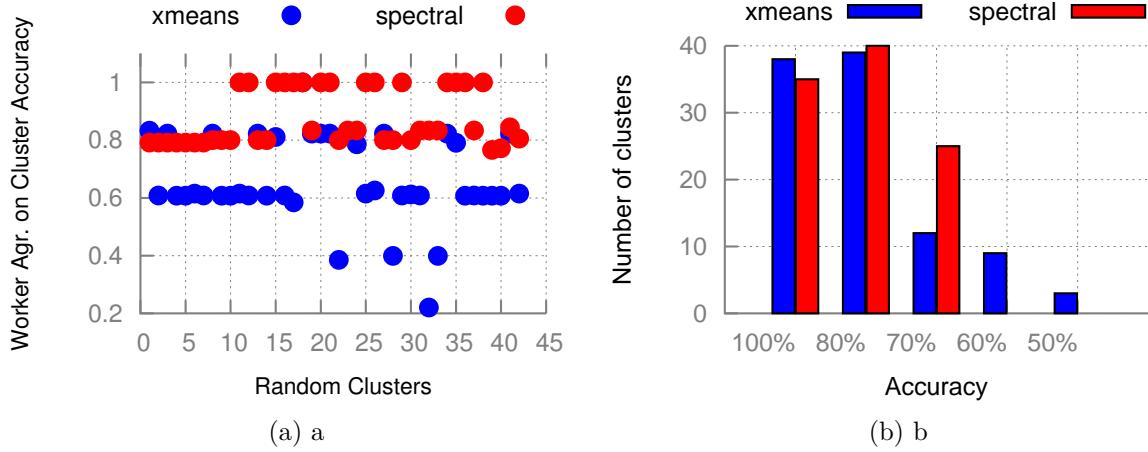


Figure 15: (a) Worker agreement on cluster accuracy for *spectral* and *x-means* clustering. (b) Cluster accuracy for the *spectral* and *x-means* clustering approaches.

Entity Retrieval Evaluation

Figure 16a presents a detailed comparison between the $P@k$ for the different methods. The proposed approaches outperform the baseline and state of the art at all ranks. The precision is highest at $P@1 = 0.6$ whereas for the later ranks it stabilizes at 0.4. In contrast to our approach, the performance of the baseline and the state of the art is more uniform, and is around $P@k = 0.25$. The best overall performing approach is the retrieval approach based on spectral clustering *SP*. Table 2 shows the details about the performance of the respective approaches as measured for our evaluation metrics.

An interesting observation is that for our approaches the best performance is achieved when querying for the field *title*. In the case of the baseline, the best performance is achieved when querying for the field *body* (B_b) while the same is inconclusive in case of the state-of-the-art methods ($S1_t$ and $S1_b$). We achieve a significantly higher retrieval performance when using the title field. This can be explained by the fact that entities that match a query on their *title* field when compared to those that match a query on their *body* field, have a higher likelihood of being an exact match.

The high gain in performance through our methods (*SP* and *XM*) stems mainly from the two steps in our approach. The first step expands the result set with relevant entities as shown in Figure 16b. The figure shows the number of relevant entities corresponding to the different grading scales as described in Section 6.3. In all cases we note that our methods find more relevant entities. The second step which re-ranks the expanded result set helps

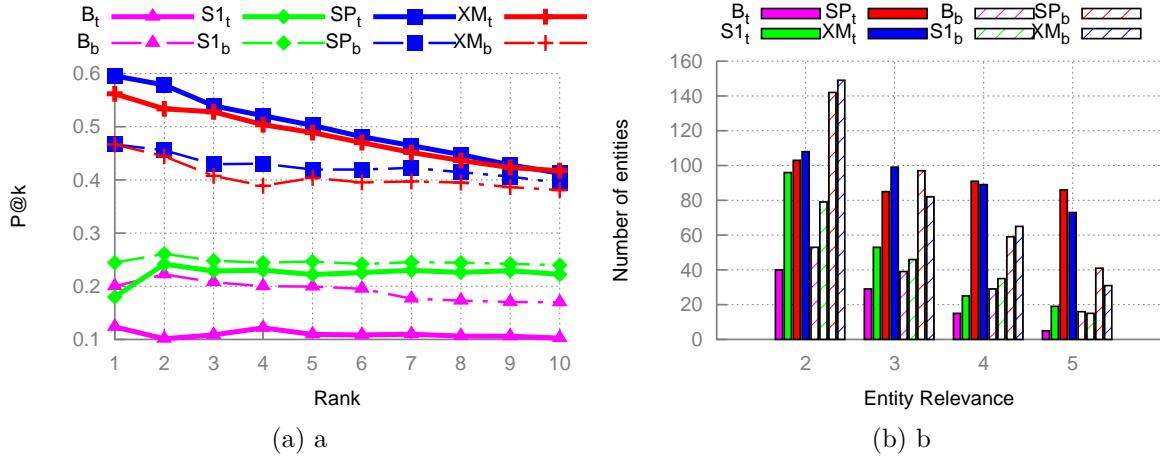


Figure 16: (a) P@k for the different entity retrieval approaches under comparison. (b) The relevant entity frequency based on their graded relevance (from 2-Slightly Relevant to 5-Highly Relevant) for the different methods.

in reducing the number of ‘*non-relevant*’ entities. We find that $S1_t$ has a 14% decrease of non-relevant entities, whereas SP_t and XM_t depict a 35% decrease, respectively. In second case where we query the *body* field, the number of ‘*non-relevant*’ entities for $S1_b$ decreases by about 13%, while SP_b and XM_b depict a 24% decrease.

	B_t	B_b	$S1_t$	$S1_b$	SP_t	SP_b	XM_t	XM_b
P@10	0.103	0.170	0.222	0.240	0.413	0.394	0.417	0.381
R@10	0.052	0.089	0.112	0.118	0.206	0.219	0.216	0.215
MAP	0.110	0.191	0.224	0.246	0.497	0.426	0.482	0.407
Avg(R)	0.031	0.058	0.063	0.074	0.132	0.133	0.131	0.130

Table 2: Performance of the different entity retrieval approaches. In all cases our approaches are significantly better in terms of P/R ($p < 0.05$ measured for *t-test*) compared to *baseline* and *state of the art*. There is no significant difference between *SP* and *XM* approaches.

We additionally analyze the performance of the entity retrieval approaches through the *NDCG@k* metric. Figure 17 shows the NDCG scores. Similar to our findings for *P@k* presented in Table 1, our approaches perform best for the query field *title* and significantly outperform the approaches under comparison.

Next, we present observations concerning the different *query types* and the entity result set expansion (see Section ??) parameters. In Figure 18a we show the improvement we gain in terms of MAP for the different *query types*. We observe that there is quite a variance for

the different query types, however, in nearly all cases, the biggest improvement is achieved through the *SP* approach. Interestingly for the query type ‘*Creative Work*’ the state of the art is nearly as good as the *XM* approach, whereas in the case of ‘*Weapon*’ the baseline performs best. One possible explanation for this is that in the case of ‘*Creative Work*’ the explicit entity similarity statements are abundant.

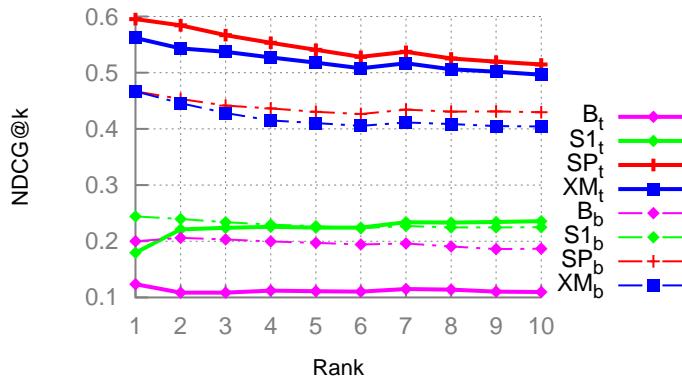


Figure 17: NDCG@ k for B_1 , $S1$ and SP , XM

Addressing the case of optimizing our retrieval approaches, *SP* and *XM*, we experimentally show the impact that the expansion of the result set has on the measured performance metrics. Here, we show the impact on the *average NDCG* score. Figure 18b shows the performance at average NDCG for the varying *cluster size* and *number of entities* added (result set expansion) for every entity in E_b . The best performance is achieved for a rather smaller *cluster size* ranging between 5 and 10 entities per cluster. Regarding the number of entities with which the result set is expanded for every e_b , the best performance is achieved by expanding with one entity per cluster. The increase in cluster size and number of entities attributes to a decrease in performance.

6.4 IFC - Point cloud schema

As part of WP3, a schema extension that allows an integration of point cloud data sets as a generic geometrical representation into the Industry Foundation Classes (IFC) schema was being created which has been extensively documented in D3.5.

Aspects evaluated

As with any standardization effort, the most *valuable* evaluation are based on:

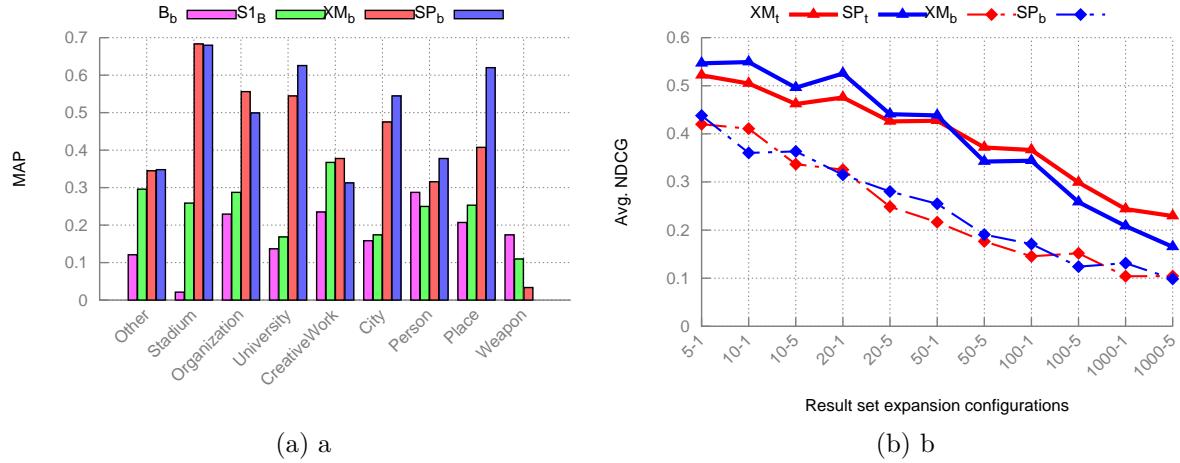


Figure 18: (a) The aggregated MAP for different query types and for the different retrieval approaches (note, we show the results for field *body* where baseline performs best). (b) The various configurations for the number of expanded entities for *SP* and *XM*.

1. the **quality aspects of the schema** in terms of *syntactical* and *semantic* correctness
2. the **feasibility aspects** of the schema in terms of stakeholder needs, information exchange requirements and external modelling experts informal plausibility checks.
3. the **adoption aspects** rate of the standard in software and industry applications.
An interoperability standard is only valuable if used and adopted in practice.

Given the short time (D3.5 was finished in M30), the adoption rate could *not* be evaluated in the scope of the project.

Method

Quality, syntactical correctness The ISO 10303-11 (EXPRESS) schema definition was developed using clear text editors. To prove its correctness it has been parsed and validated with the well-established and proven tools JSDAI³⁵ by LKSoftWare GmbH and the STEPCode³⁶ tool `schema_scanner`, as well as by a customized version of the IFCOpenShell³⁷ toolkit (which has also been used in subsequent prototypes described and evaluated here).

³⁵<http://www.jsdai.net/>

³⁶by the American National Institute for Standardization NIST stepcode.org

³⁷<http://ifcopenshell.org>

Quality, semantic correctness From a semantic point of view basic modelling constructs, naming conventions and modeling guidelines of the overarching IFC schema have been followed very closely.

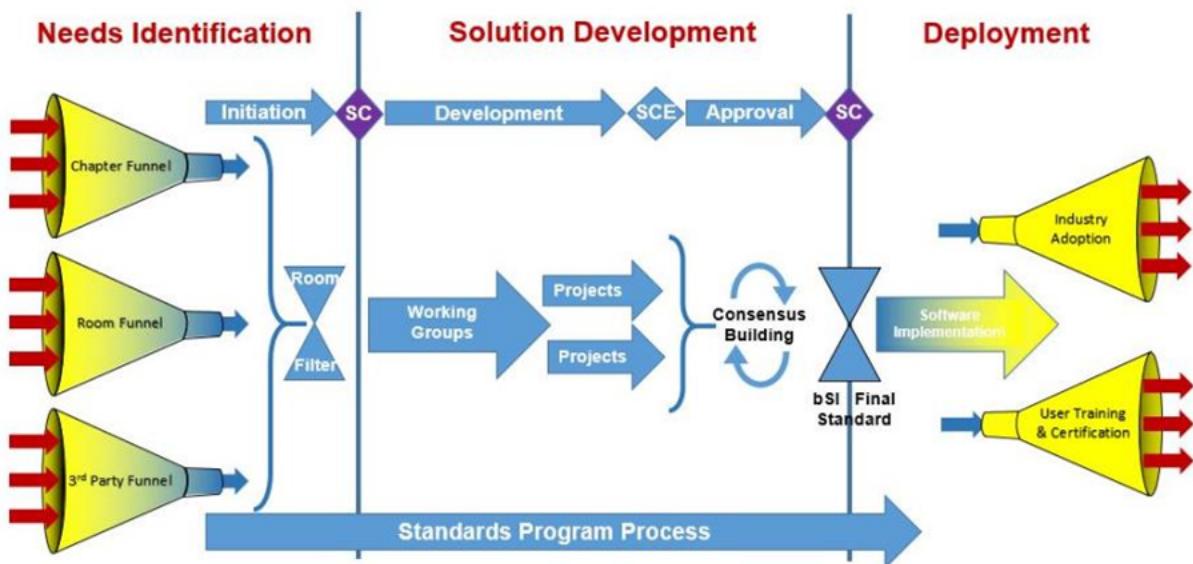


Figure 19: The standardization pipeline of the buildingSMART organization. The IFC Point Cloud schema has been introduced into the pipeline in October 2015, after the idea has been embraced in 2012

Feasibility The schema has been presented to both experts from the IFC standardization community and experts in the field of interoperability as well as to experts from the computer graphics domain. Furthermore, the resulting schema has been checked against the requirements resulting from WP2 Deliverables and end-user whishes gathered in WP7. The schema and its documentation has been presented to the buildingSMART community during the World Summit in Singapore in October 2015 as a first step into the standardization pipeline which is illustrated in figure 19. The results have been presented to the scientific community during the EG-ICE workshop in 2015 and informal requests for comments have been send out to individual experts in the field that work with both point cloud data sets as well as with IFC models. Individual persons include Thomas Liebich³⁸, the head of the Modeling Support Group (MSG)³⁹, Prof. Robert Amor (Aukland, NZ) and Prof. Ioannis Brillakis (Cambridge University).

³⁸also a member of the advisory board of the DURAARK project

³⁹the core technical group responsible for the development and maintenance of the IFC schema

Datasets and Evaluation Activities

The correctness, feasibility and performance of the schema has been tested and validated by instance population datasets from the DURAARK repository. As a main source, the Haus 30 dataset (see Figure 20) which is available in both representations (explicit IFC model and point cloud scans) provided by an industry stakeholder has been used as an example. All forms of associations of point clouds to explicit geometry provided by the schema have been covered in a variety of configuration permutations along with their resulting compression mechanisms. The resulting datasets are available for reproducible evaluation by third parties both in clear-text encoded STEP Physical File Format (SPFF / p21) as well as their counterparts serialized as HDF5 (see section 6.5).

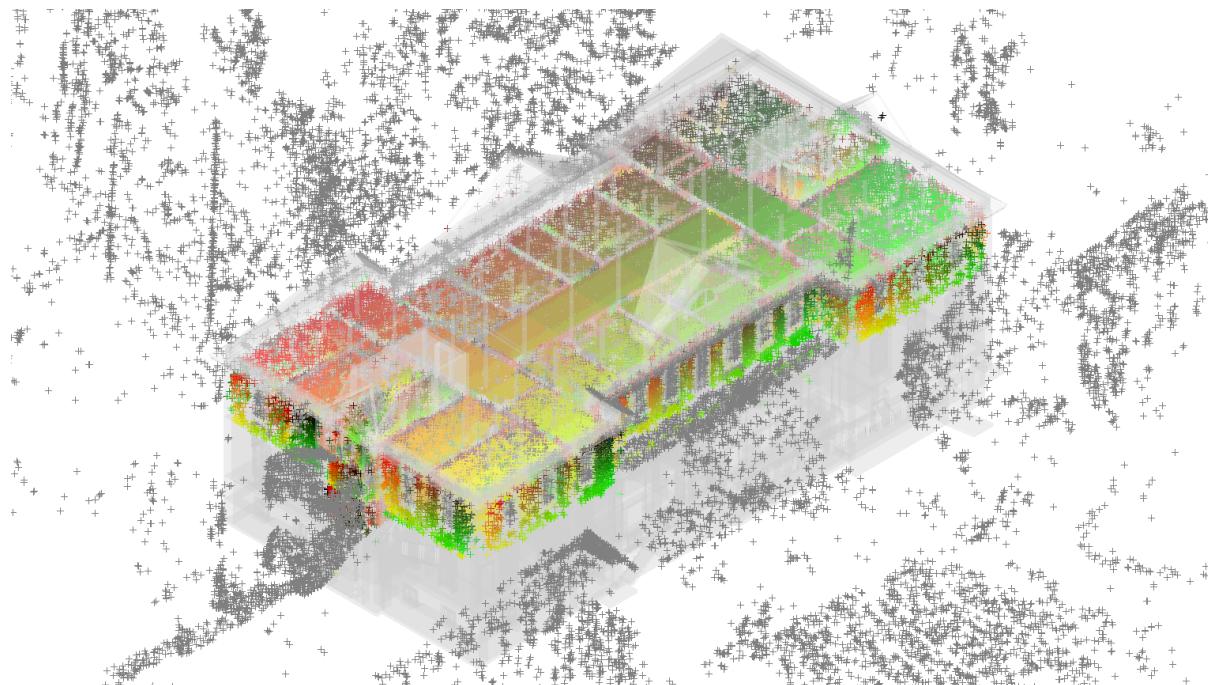


Figure 20: Screenshot of a viewer application used for the validation of instance data sets of the Point Cloud extension schema. The depicted model is the Haus 30 data set available in the DURAARK repository

The building model consists of 364543 entity instances describing a total of 819 building elements and is enriched with a point cloud of 13 million points. These point clouds are later subdivided over 3283 associated segments. For further performance metrics regarding the compression rates, object metrics and files sizes please refer to section 6.5.

Results

The technical schema validation in all used tools resulted in no errors. The schema is thus feasible for implementation in its current form in commercial applications such as CAD/BIM packages, downstream processing applications, scientific prototypes etc. Future applications and the inclusion of further data sets is already ongoing and a perspective adaptation by the buildingSMART organisation is actively undertaken. The general perception of the schema from a modelling point by both experts and industry practitioners has been generally positive. Concerns and suggestions were generally focused on the overall approach to keep data of inherently different nature and with (today) separated software eco-systems separated.

6.5 IFC HDF5 Serialization

As a part of D3.5 a serialization format was developed and suggested to efficiently store and process point cloud data using the suggested IFC schema extension (see section 6.4)

Aspects evaluated

The following aspects have been evaluated:

1. Compatibility and syntactic correctness HDF5
2. File size and compression rates

Method

In order to ensure the correctness of the generated HDF5 files of both current IFC files in versions 2x3 and 4, as well as the files containing point cloud data, the reference implementations of the HDFGroup have been used to generate and process the data. In particular the HDFView⁴⁰ application along with its underlying libraries has been used to inspect, debug and navigate the resulting data sets. The developed viewer application (see Figure 20) also make use of the reference implementation libraries.

⁴⁰<http://www.hdfgroup.org/products/java/index.html>

Datasets and Evaluation Activities

All performance test have been executed using the Haus 30 dataset available from the DURAARK repository. Different combinations point cloud registration- and compression-modes have been tested that are documented in the following section and the D3.5 Deliverable.

Results

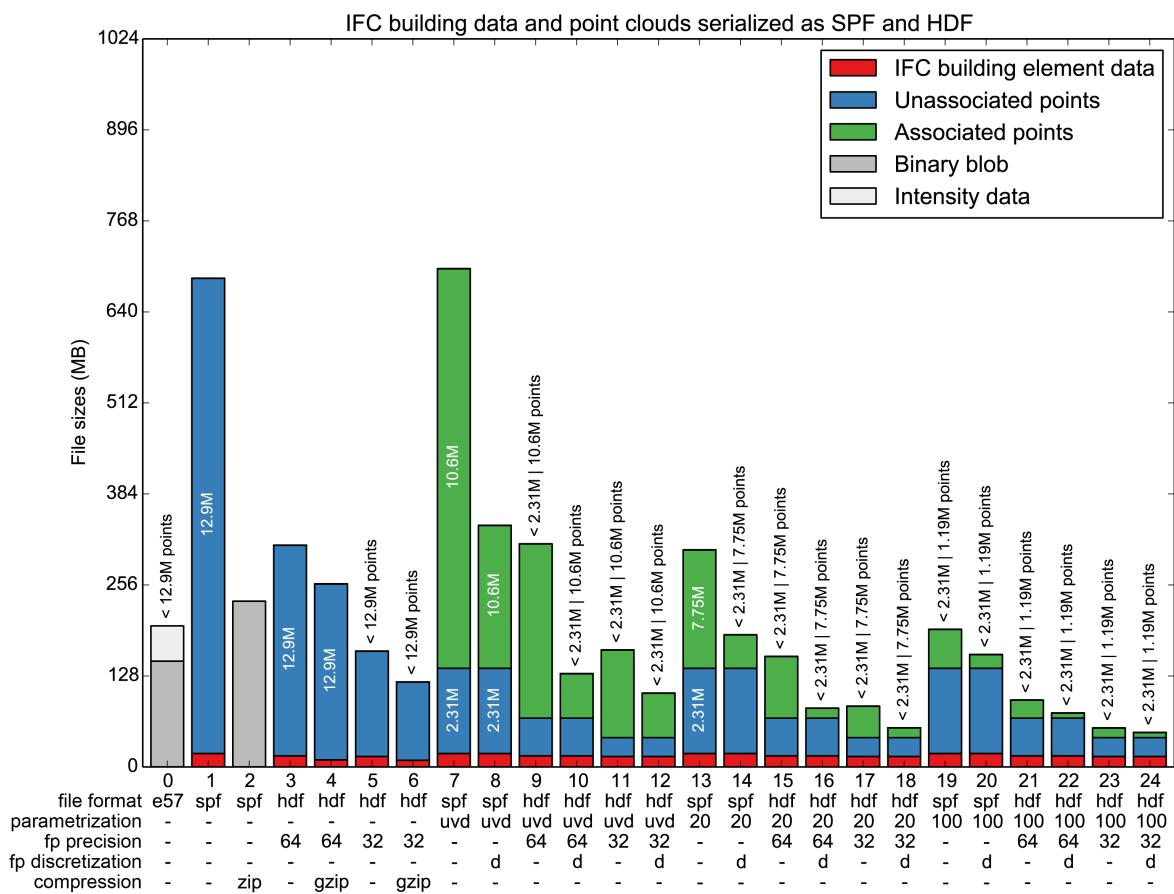


Figure 21: Chart of file sizes achieved with different combinations of HDF5 vs. SPFF serializations and compression strategies

The results of the file size and compression rate combinations are illustrated in figure 21. Contributions to file size are subdivided according to *a*) original IFC building element data *b*) unassociated point cloud segments, i.e. subsets that do not conform to original building element surfaces and *c*) associated points, i.e. subsets that are sufficiently close

to building element surfaces in order to be parametrized. Columns 0 and 1 show the file sizes of the original e57 and IFC SPF serializations, a naive zip compression is added in (3). Even though the file size of naive HDF5 serialization (column 4) is larger than the zip compressed SPF, by means of its block-level compression and hierarchical nature, HDF5 allows rapid random access as opposed to SPF which has to be read and parsed into the main memory entirely before being able to access data items.

Note that the e57 file (column 0), which has been the source document for the point clouds, is comparable in size to the 32bit uncompressed HDF5 serialization (5) when the non-Cartesian point information is deducted. This is to be expected as both carry the same number of points in the same 32bit floating point format. This illustrates that the HDF5 file format itself does not introduce any significant inefficiencies in terms of how the point cloud data is stored.

The introduction of associations between point cloud segments and building elements comes with a minimal cost, as association entities need to be populated. This can be seen when comparing cases (3) and (9). However, as discussed in DURAARK Deliverable D3.5, the parametrization onto bounded surfaces allows the use of an integer discretized notation for the parametrized components. Combined with a precision loss under typical scanner's resolution, this yields a tremendous file size reduction in both SPF (7-8) and HDF (9-10). At this point the HDF file (10) is smaller than the original e57 file, carrying the same number of points, enriched with the complete IFC building model instantiation. Higher compression ratios can be obtained by introducing lossy grid quantization by subsampling point clouds onto grid structures defined in the parametric coordinate space of the building element surfaces.

7 Evaluation - Geometric Enrichment Components

In this Section we will outline our evaluation of the geometric enrichment tools developed primarily in WP4 and WP5. The general focus of these geometric enrichment tools is to bridge the gap between the metadata-rich BIM models given as IFC representations and the purely geometric Point Cloud representations given as E57 files. While the former are specifically designed to include any potential valuable architectural data, the latter even lacks basic topological information (e.g. which points belong to the same object). *Geometric Enrichment* here now means to (heuristically) derive relevant information such as the aforementioned topological structure from the purely geometric information given in the Point Cloud data. In addition to this heuristic derivation, a lot of information can be extracted by *combining* both representations of the same building in order to e.g. analyze the structural differences between a BIM model generated in the planning phase of a building and the actual building as measured using a Point Cloud scanning method.

7.1 Difference Detection

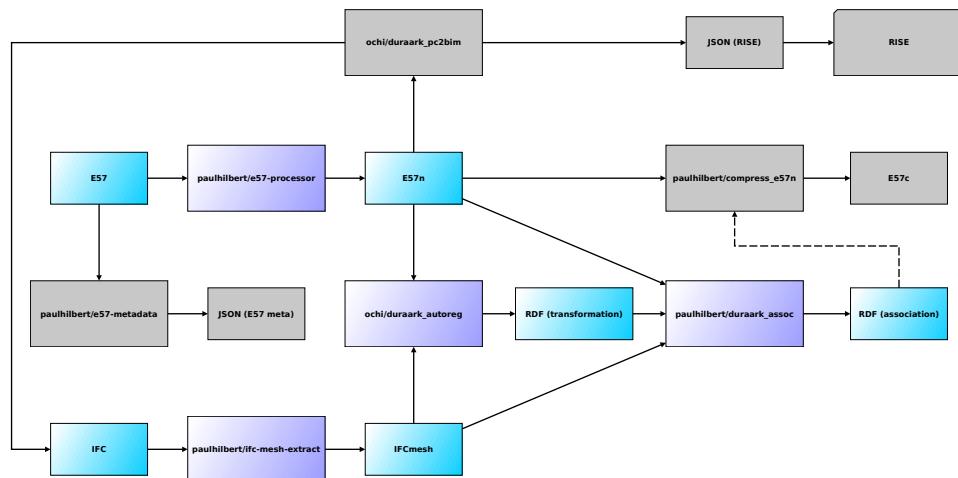


Figure 22: Microservices used for Association / Difference Detection.

The **difference detection** is a task of comparing different datasets in respect to the building elements which they consist of. This chapter presents the two implementations of the difference detection toolset developed as a part of DURAARK workbench:

- Implementation in the DURAARK WorkbenchUI

- Implementation within a Stakeholder Work Environment

Both implementations use the same **registration** and **association** tools to obtain the necessary information to perform the **difference detection**. The visualization of the results of the first implementation is done using Potree, while the second one is using the Volvox environment integrated within the Rhino/Grasshopper application.

The focus of the evaluation activities is set on the Volvox toolset, which enables to precisely measure relationships between Point Clouds and BIM geometry. Additionally it is possible to create necessary tools on-the-fly with the Grasshopper scripting components.

Difference Detection with DURAARK components

The Difference Detection is a task accomplished by the association toolset developed in WP 4 (see fig. 22). This task is tackled by solving the complementary problem: Detecting similarities between two representations of architectural geometry.

In order for this to work both representations need to be geometrically aligned which is accomplished by the **registration** tool which computes the transformation aligning the first to the second representation. Given this registration the **association** tool now associates *subsets of points* to either *subsets of points* or *IFC building elements* depending on whether the tool is used for point-cloud-to-point-cloud or point-cloud-to-IFC association. The output of the latter tool is a RDF description of associated, geometric subsets. For example the following RDF snippet describes a set of point indices (“rel:pointSubsetContains”) in a Point Cloud (“rel:pointSubsetOf”) which is associated to an IFC building element represented by its globally unique identifier (“rel:subsetRepOf”):

```
<PC_Rislokka_upper>
<rel:globalUniqueId> "{13af2df2-5a16-400e-90f5}"^^<xsd:string> .

<SBS_PC_Rislokka_upper_10116165214659918227>
<rel:globalUniqueId> "SBS_{13af2df2-5a16-400e-90f5}_101161652"^^<xsd:string> ;
<rel:pointSubsetContains> "224-229 331-454 456 458"^^<types:indexSet> ;
<rel:pointSubsetOf> <PC_Rislokka_upper> ;
<rel:possibleScanHits> "906"^^<xsd:nonNegativeInteger> ;
<rel:subsetRepOf> "10RXOPRpjA$ee5bm0a7zft"^^<xsd:string> .
```

Note that this representation implicitly encodes the differences between both datasets, since

- IFC elements with none or too few points indicate objects not scanned, and

- points not associated to any elements are likely to represent objects not present in the IFC model.

For example see fig. 23 for a simple visualization of points *not* part of any association.

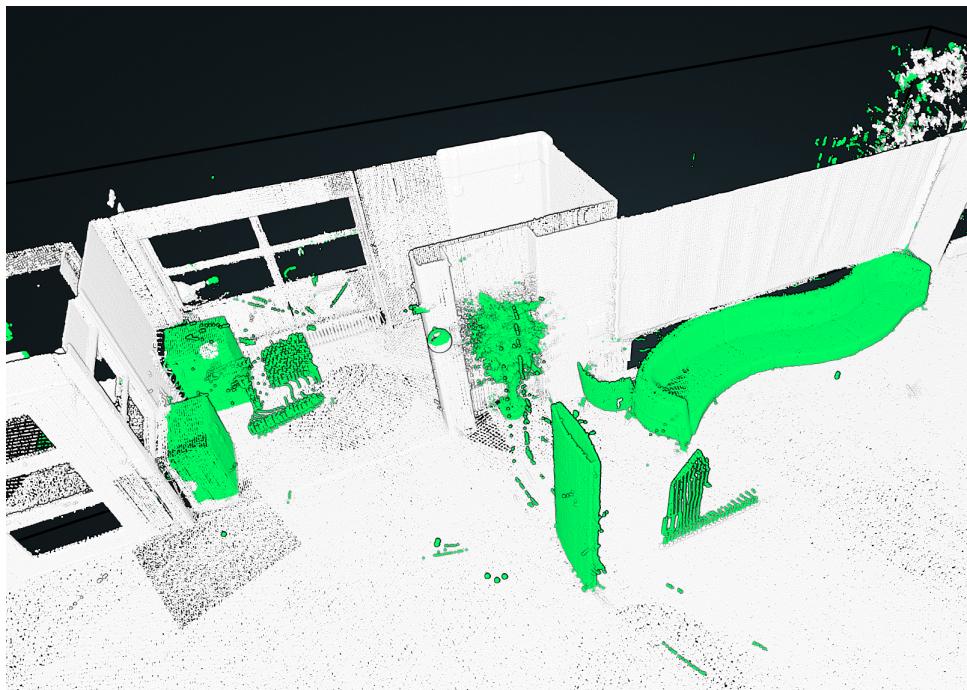


Figure 23: Visualization of points not associated to any IFC building entities.

Difference Detection for the WorkbenchUI As mentioned above the output of the association tool itself is in a machine readable format for further processing (for example for the compression tool). However in order to showcase the deviation and to some extent evolution of a building, an easily navigatable visualization is necessary. For the DURAARK framework this is accomplished via three additional tools:

- The **Difference Detection** component which uses the association output in order to generate an annotated E57 Point Cloud.
- The modified **PotreeConverter** generally used in the DURAARK framework for creating WebGL-based, online navigatable Point Cloud visualizations including this annotated variant.
- The **Volvox extension** developed to precisely measure the registration/association errors and visualize and compare the association results.

The **Difference Detection** and **PotreeConverter** tools were used to render the visualization shown in fig. 23.

Difference Detection in Stakeholder Work Environment

We've evaluated the Difference Detection tool according to the WP4 description, which assigns specific tasks required:

- registration
- transfer of structure and semantics
- identification of inconsistencies

In the following chapter these are named respectively:

- registration
- association
- difference detection of the 3d scan and BIM model

The **registration** tool is meant to robustly register the points clouds against the geometry of the corresponding BIM model. The tool creates a transformation matrix which when applied to the Point Cloud transforms it to comply with the BIM geometry. As a result, the user obtains an RDF file which stores the transformation matrix and information on the geometry it should be applied to.

The **association** tool is capable to transfer information from the BIM model to the Point Cloud, and annotate different portions of the Point Cloud in respect to the parts of the BIM geometry. The product of this tool is an RDF file with a list of indices assigned to identifiers (UUID) from the BIM model.

To execute the **difference detection**, both of the tools described above have to be used to generate appropriate RDF files. For the evaluation purposes, we've developed a Volvox extension which is able to visualize, compare and edit the associated Point Clouds. We've used this toolset to perform the precise evaluation described in the following section.

Aspects evaluated

To perform the difference detection, a proper preparation of data is required. Given a BIM model and a 3d scan, user has to use the developed tools in the following order: registration, association (to obtain transformation and association RDF files) and finally difference detection, which is based on information from the RDF files. We measure registration and association precision using qualitative statistical methods (average and standard deviation), and the difference detection precision is measured quantitatively.

Parts of the evaluation:

1. Precision of the registration
2. Precision of the association
3. Precision of the difference detection
4. Utility: what is the effort to implement the tool?
5. Relevance. is the data useful for the given use cases and stakeholders?

Method

The evaluation of the Difference Detection component took place on several levels. Each of them used a dataset with existing 'ground truth' data, against which the results could be compared.

1. The precision of **registration** is evaluated by measuring distances of a set of 1000 random points from a Point Cloud to relevant parts of the BIM model. Average and standard deviation are used, most characteristic data plots (histograms) are provided. As the BIM models are in most cases a simplified geometrical representation of the buildings, we account the measurement error resulting from the missing architectural details.
2. The evaluation of **precision of association** is based on a comparison between IFC entity and relevant Point Cloud. Both average and standard deviation are measured. Additionally visual inspection of histograms was performed. Some of the most characteristic histograms are provided below as a commentary. In terms of the evaluation of precision, we account the error resulting from the simplified BIM

models, moreover we account for the error resulting from the ε association threshold value used in the association process. This ε value is basically the maximum distance that the point under consideration may have from the location where the simulated laser ray hit the surface in the IFC model. Please see Section 4.3 in D4.2 for more details.

3. The precision of **difference detection** is evaluated by a manual comparison between known changes in a building and those detected with algorithm. A perfect result for difference detection is to give an information on all the changes in the history of the building, without detecting false positives.
4. The evaluation of utility is based on our implementation of the tools into the Volvox plug-in
5. Relevance: Deliverable D2.2.1 defined several use cases for this DURAARK component, which are relevant to stakeholders (UC4, UC5, UC6). If these are met, it implies the tool has relevance to stakeholders.

Datasets and Evaluation Activities

The evaluation has been conducted on 3 datasets, which have been selected from the DURAARK data repository. They have been selected, as all of them show explicitly a difference over time or have a known deviation between a 3d scan and BIM geometry.

1. CITA Byg72 - three different stages of renovation process and one BIM model as the 'ground truth'
2. Plan3D Haus30 - one 3d scan and one BIM model as 'ground truth'
3. Statsbyg Rislokka - one 3d scan and one BIM model as 'ground truth'

To evaluate the datasets, we've performed the registration and association of the Point Clouds. With the data from the RDF files we were able to extend the Volvox plug-in's functionality to find and interpret the differences in various configurations. Particularly the data from the association file proved to be useful to accomplish the difference detection, as the RDF files provided the information on theoretical amount of points (given the 3d scan meets certain registration quality). By comparing the theoretical amount of points with the actual Point Cloud count we were able to deduct the missing parts of the building.

To detect the new parts of the building in a 3d scan, an IFC reconstruction should be performed which would provide a set of geometries to associate with the Point Cloud.

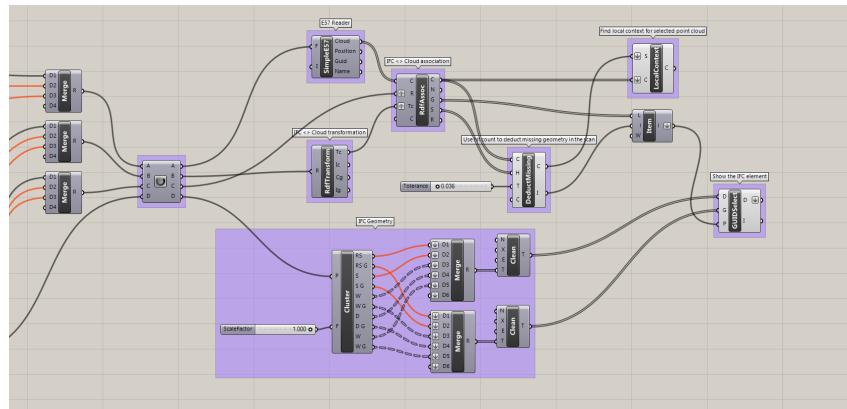


Figure 24: Basic evaluation setup.

To read the geometry from the IFC model in Grasshopper, we explored various tools (Solibri Model Viewer⁴¹, VisualAqr⁴², Geometry Gym⁴³). All of them showed inconsistent results. Because of the faulty importers/exporters, we were not able to completely automate the evaluation process, therefore some visual inspection was required - accompanying the histograms.

Results

While the registration and association tools are automated, we were forced to perform the difference detection within Volvox in a semi-automated way, due to the problems in reading IFC geometry in Grasshopper correctly. The results of the difference detection proved nevertheless to be unequivocal in terms of interpretation.

We've performed a detailed evaluation of the 3 tools based on 3 datasets:

CITA Byg72

⁴¹<http://www.solibri.com/products/solibri-model-viewer/>

42 www.visualarg.com/

⁴³ <https://geometrygym.wordpress.com/>

1. Registration: Transformation of the first scan has an error of 150 mm, therefore we use the 3rd (newest) scan to analyse the registration, which has an average error of approximately 50 mm. The error mostly comes from the fuzzy association which accounts some small parts of furniture as building geometry (it also happens in the best case association pair).
2. Association: we can't compare association for the oldest scan because of the registration error (approx. 150 mm). We use the newest one for evaluation, which has some minor topological differences compared with the IFC model (which were detected with the difference detection). The best association pair has an average distance to the Point Cloud equal to 58.9 mm. This value is greatly influenced by the association of some of the furniture with the building geometry. Additionally the association errors are sporadically caused by faulty results of the IFC importer, which mismatches especially the wall thicknesses.
3. Difference detection: There are 2 changes in the building geometry over time, which are 2 walls removed during the refurbishment. We've automatically detected 3 changes among 3 scans and from those changes one is erroneous. The detection error is caused by imprecision of the 3d scan, and it is easily distinguishable from the properly detected changes. The other 2 changes are clear and easy to identify. The data used to identify changes is the potential hit count - thanks to those values we can compare actual Point Cloud count and calculate proportion between that number and theoretical hit count value. It became a convenient way to detect changes.

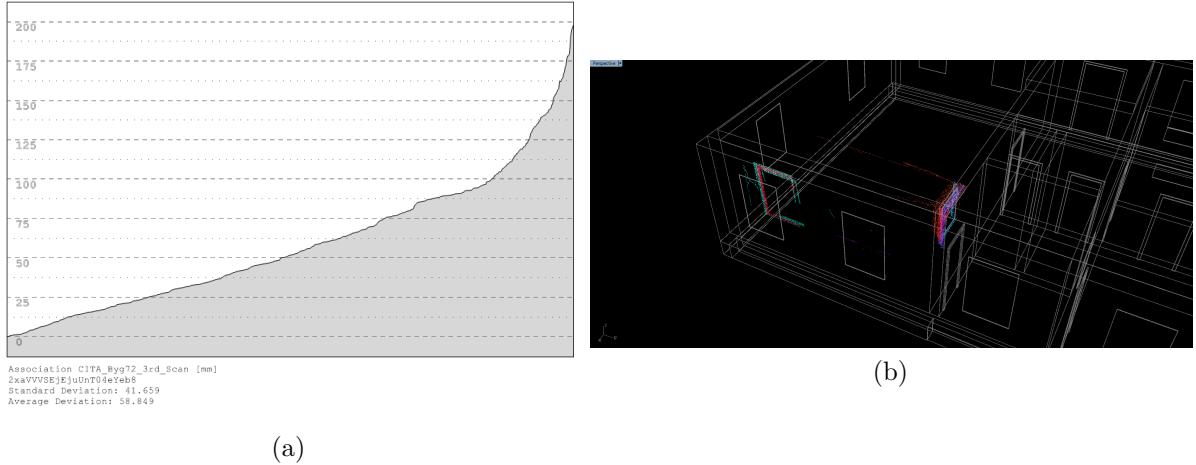


Figure 25: (a) Best association pair for CITA Byg 72 [mm], (b) Incorrectly detected change in the Byg 72, caused by the lack of data in the 3d scan.

Plan3D Haus30

1. Registration: We've used only the wall geometry to analyse the registration precision because of the faulty IFC importer. The average mesh to Point Cloud distance is below 32 mm. This value is greatly affected by the lack of architectural details in the IFC model. Please refer to the association part.



Figure 26: (a) Registration overview, (b) Registration histogram [mm].

2. Association: We are unable to evaluate the worst case association pair, because of

the before mentioned IFC importer inconsistency. Contrary to that, we've evaluated the best case pair, which is by average as close as 10 mm from the IFC geometry. From the illustration 27a and the histogram 27b, it is visible that the error value is greatly influenced by the architectural detail not being a part of the IFC model yet is being included because of the ε association threshold value.

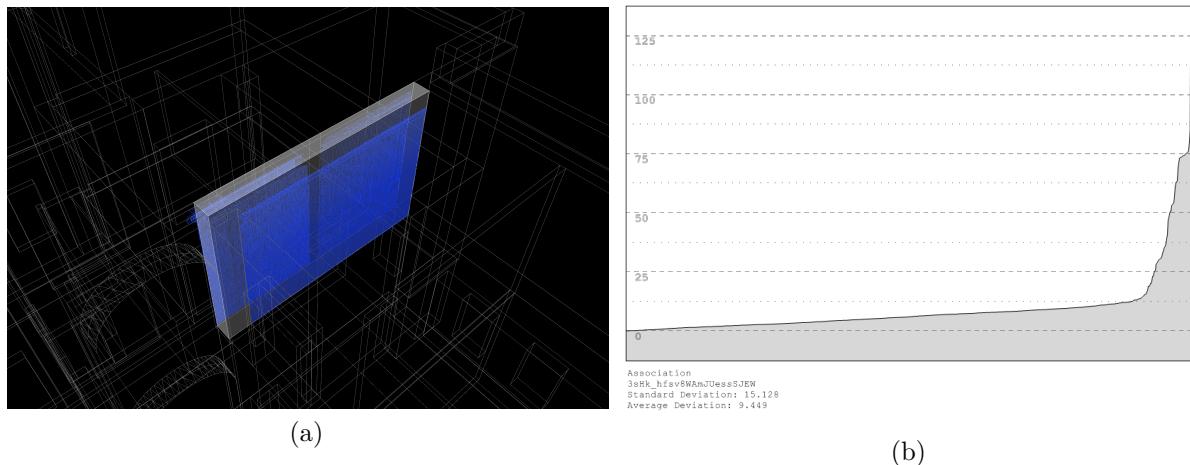


Figure 27: (a) Best association case for Plan3D, (b) Histogram of the best association pair [mm].

3. Difference detection: We've analysed the provided scan of one floor of the building which was compared with IFC model of the whole building. We've correctly detected one change between the scan and the provided model.

Statsbyg Rislokka

1. Registration: The average registration precision measured is below 32 mm. From the histogram 28 we can read that there are few outliers associated incorrectly as seen in the picture 31.

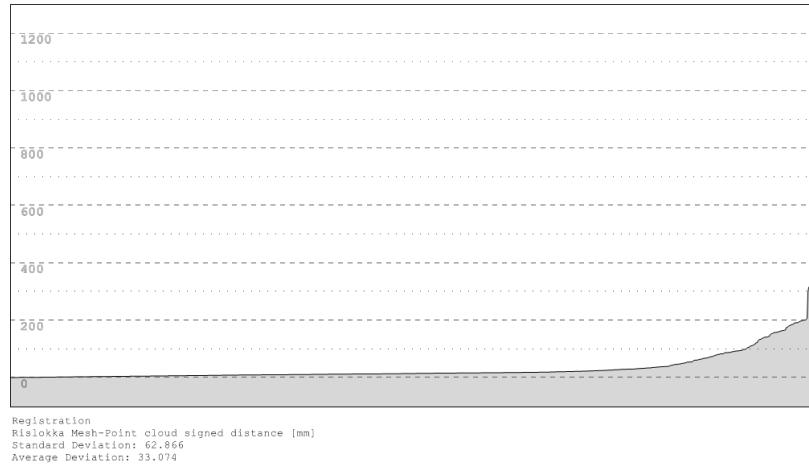


Figure 28: Rislokka registration histogram.

- Association: There are few outliers incorrectly associated with some IFC geometry, which are easy to spot in the overview picture 31. The best case association pair has an average error of less than 5 mm. Again we are not able evaluate the worst case matching because of the IFC inconsistency.

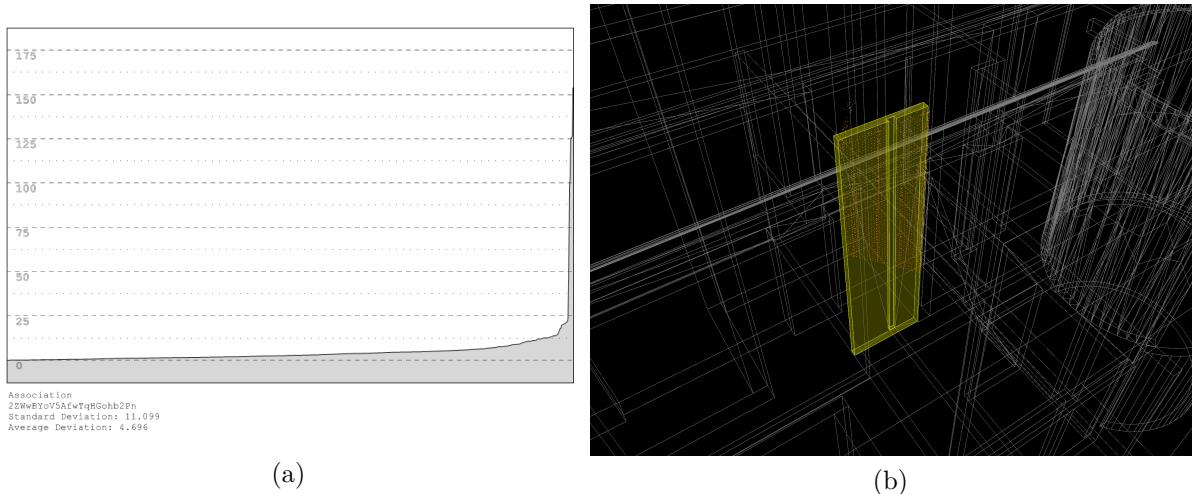


Figure 29: (a) Best case association histogram, (b) Best case association pair.

- Difference detection: The evaluation of the difference detection in the Rislokka building showed that the potential hit count measure doesn't take into account scanning errors inherited from the scanning device - that is - 3d laser scanners can't

properly scan transparent materials such as glass, therefore some windows were detected as missing (low Point Cloud count compared with the high potential hit count) 30.

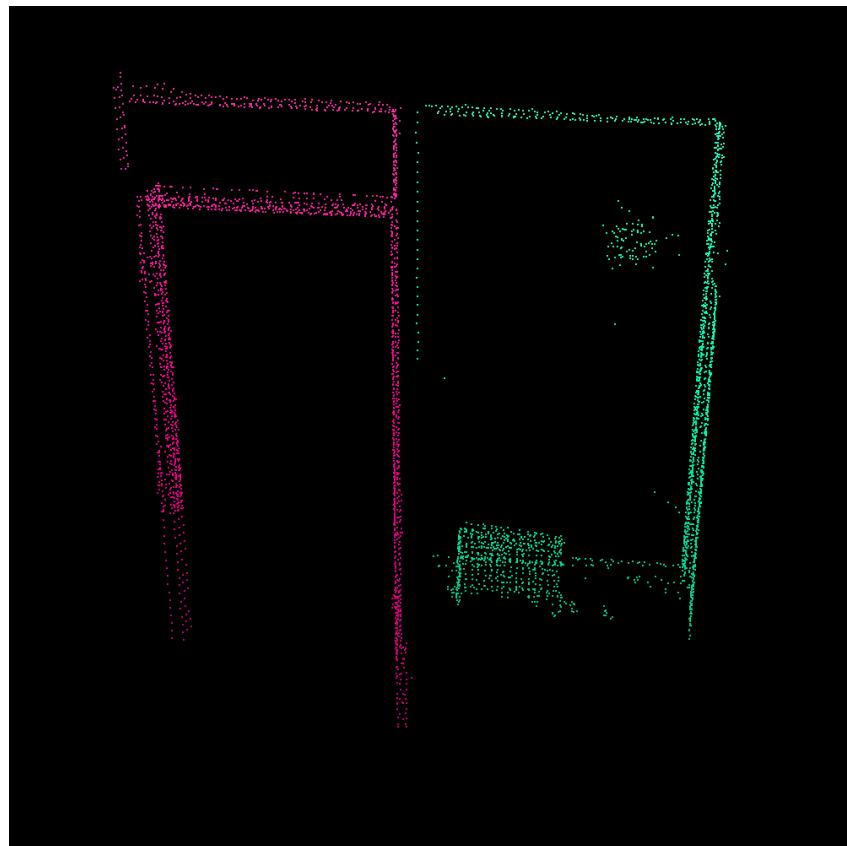


Figure 30: The parameter "Potential hit count" for this window is 65000. It should show whether an object exists or not. It is however not too reliable, as it is i.e. confused by transparent materials such as glass.

Summary

Both registration and association tools proved to work in our evaluation cases. With some effort set into the interpretation of the data provided with RDF files, we were able to perform a difference detection within the Volvox environment. While the difference detection was automated, a visual inspection of the differences was required in order to identify false positive detections. Those were mostly caused by low quality of the 3d scans or association mismatches resulting from transparent materials.

It seems that the association part would perform much better with some outlier removal

post-processing, as displayed in the Rislokka dataset 31. Potentially a great part of outliers could be removed automatically, using basic statistical measures (standard deviation etc.)

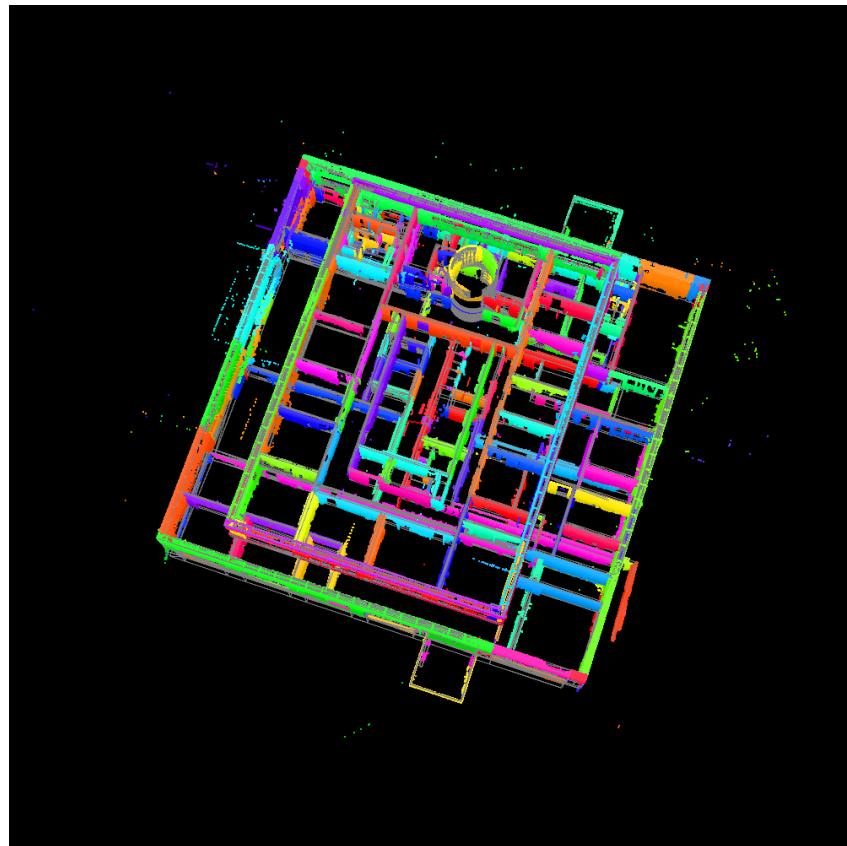


Figure 31: Rislokka overview with incorrectly associated outliers.

DataSet	Registration error
CITA Byg72	50 mm
Plan3D Haus30	32 mm
Statsbyg Rislokka	32 mm

DataSet	Association best case average distance
CITA Byg72	58.9 mm
Plan3D Haus30	10 mm
Statsbyg Rislokka	5 mm

DataSet	Changes detected	Incorrect detections
CITA Byg72	3	1
Plan3D Haus30	1	0
Statsbyg Rislokka	0	2

7.2 IFC reconstruction

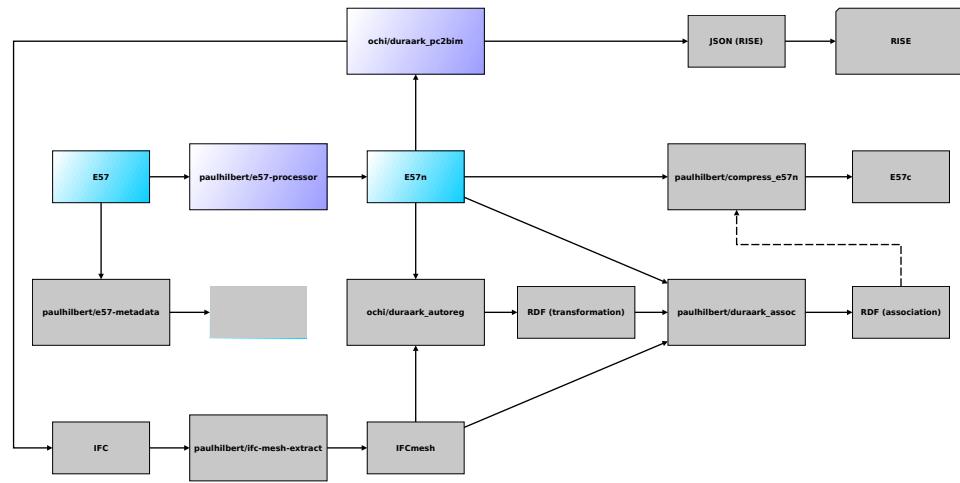


Figure 32: Microservices used for IFC reconstruction.

This section contains the evaluation of the IFC Reconstruction Component. This component consists of a single (containerized) microservice (“duraark_pc2bim”, see fig. 32) which – given a preprocessed input Point Cloud in E57 format – produces a reconstructed, parametric building model in IFC format.

In the state used for this evaluation, the tool supported reconstruction of wall structures, floors and ceilings as well as openings for doors and windows. The determination of door opening directions as well as the detection of columns were not supported when the evaluation was performed. Additionally, the tool optionally supports the output of a JSON description of roughly the same geometric entities. This serves as a means for a more simple, direct interface to the RISE components developed in WP 5 and is not intended to be used directly with existing stakeholder software (hence it is not part of this evaluation).

The primary intended use for the tool is to produce a BIM representation to be used directly in consumer software and not for further applications inside the DURAARK

framework (although its output is a suitable foundation for the RISE component). In more abstract terms the tool fills the gap introduced by legacy buildings for which BIM data has never been produced and therefore provides a simple yet important basis for e.g. retrofitting or analysis purposes.

Aspects evaluated

1. The **relevance** of the IFC Reconstruction component is evaluated against already existing approaches for the geometric enrichment of Point Clouds and the efforts in terms of time needed to do so.
2. The **precision** of the reconstructed IFC is evaluated in terms of labelling of IFC elements (walls, doors, windows) and it's dimensions in comparison with 'ground truth' models of real world data provided by stakeholders.
3. The **quality** of the reconstructed IFC is evaluated against the needs of the stakeholders in terms of completeness, overall tolerances and drawing conventions.

Method

The evaluation builds upon already conducted evaluation activities in deliverable D7.2. the results of D7.2 showed that the most significant reconstruction failures were False reconstructed and missing reconstructed walls, doors and windows. Reconstruction failures like wall height and level height were not as significant in importance and have been implemented in the IFC reconstruction tool shortly after D7.2. Wall thickness and drawing conventions, were part of the previous deliverable, where it was argued, that they are important for the implementation in stakeholder software, but rather a post-processing step to the actual IFC reconstruction component. They will therefore not be the main focus of this evaluation. The focus of this evaluation activity are false reconstructions and missing reconstructions of the currently detected architectural elements of walls, doors and windows.

In order to evaluate the **quality**, **precision** and **relevance** two overall methods have been used:

- **Comparison: Reconstructed IFC / Point Cloud** - Systematic manual comparison of reconstructed IFC elements with the Point Cloud.

- **Statistical Analysis** - Statistical analysis of deviations between reconstructed IFC and 'ground truth' IFC collected from stakeholders and saved in the DURAARK repository.

The evaluation activities have been conducted towards the needs of stakeholders from building owners and facility management. This distinction has been taken, because this stakeholder group is the one with highest potential interest in the DURAARK system, as documented in D7.2 and 8.7) -they have as well a basic demand on precision and elements to detect (mainly walls, doors, windows, areas and relations of rooms), which is appropriate to the state of the IFC reconstruction component.

We categorize the reconstructed failures hence into three categories, **Deviating**, **False** and **Missing** reconstructions and skip precise measurements of deviations - partly, because this is part of the evaluation of the difference detection (chapter 7.1), partly as it seem redundant in terms of the stakeholder group in mind. The graphic results of the manual comparison categorizations have been colour coded respectively in Blue, Red and Magenta (fig. 33).

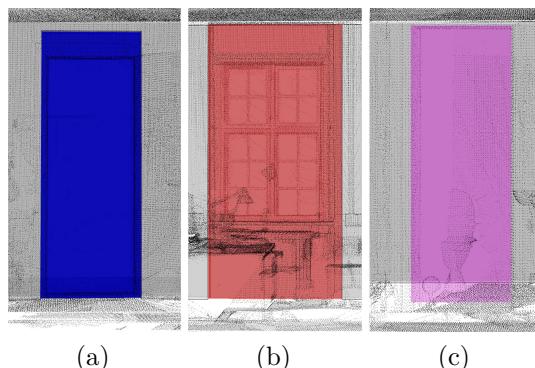


Figure 33: Categorization of the reconstructed failures into three categories, (a) **Deviating**, (b) **False** and (c) **Missing**

Besides this categorization, False reconstructions are in the description labelled as False Positive or False Negative. This means for instance for openings:

- If an opening has been detected correctly, but the reconstructed element has been labelled False, e.g. an opening for a window has been detected, but a door has been reconstructed, then the door is categorized as False Positive.

- If an opening has been falsely detected and an element has been reconstructed, e.g. a door has been reconstructed in a wall with no opening, then this is categorized False Negative.

Datasets and Evaluation Activities

DataSets The evaluation activities have been conducted on 16 datasets, from which IFC models were automatically reconstructed. The projects were selected for the evaluation activities, as they display a range of different architectural typologies, from residential, over office to open floor lofts.

CITA Byg72 Scan June 2015	LE34 Højbro Plads 3rd Floor
CITA Byg72 Scan August 2015	CITA DesignMuseum Copenhagen
CITA Byg72 Scan September 2015	Stattsbyg Rislokka middle
Plan3D Haus30	ZESO Nygade
Stattsbyg Rislokka upper	Wework 240 Bedford
Faro Apartment	Wework 255 Butler 1st Floor
LE34 Højbro Plads 1st Floor	Wework 255 Butler 2nd Floor
LE34 Højbro Plads 2nd Floor	Wework 255 Butler 3rd Floor

Table 3: Datasets used for the evaluation of the IFC reconstruction component.

Comparison: Reconstructed IFC / Point Cloud The comparison with the Point Clouds we have manually gone through all reconstructed models and annotated False and Missing reconstructions of walls; and False, Missing and Deviating reconstructions of doors and windows.

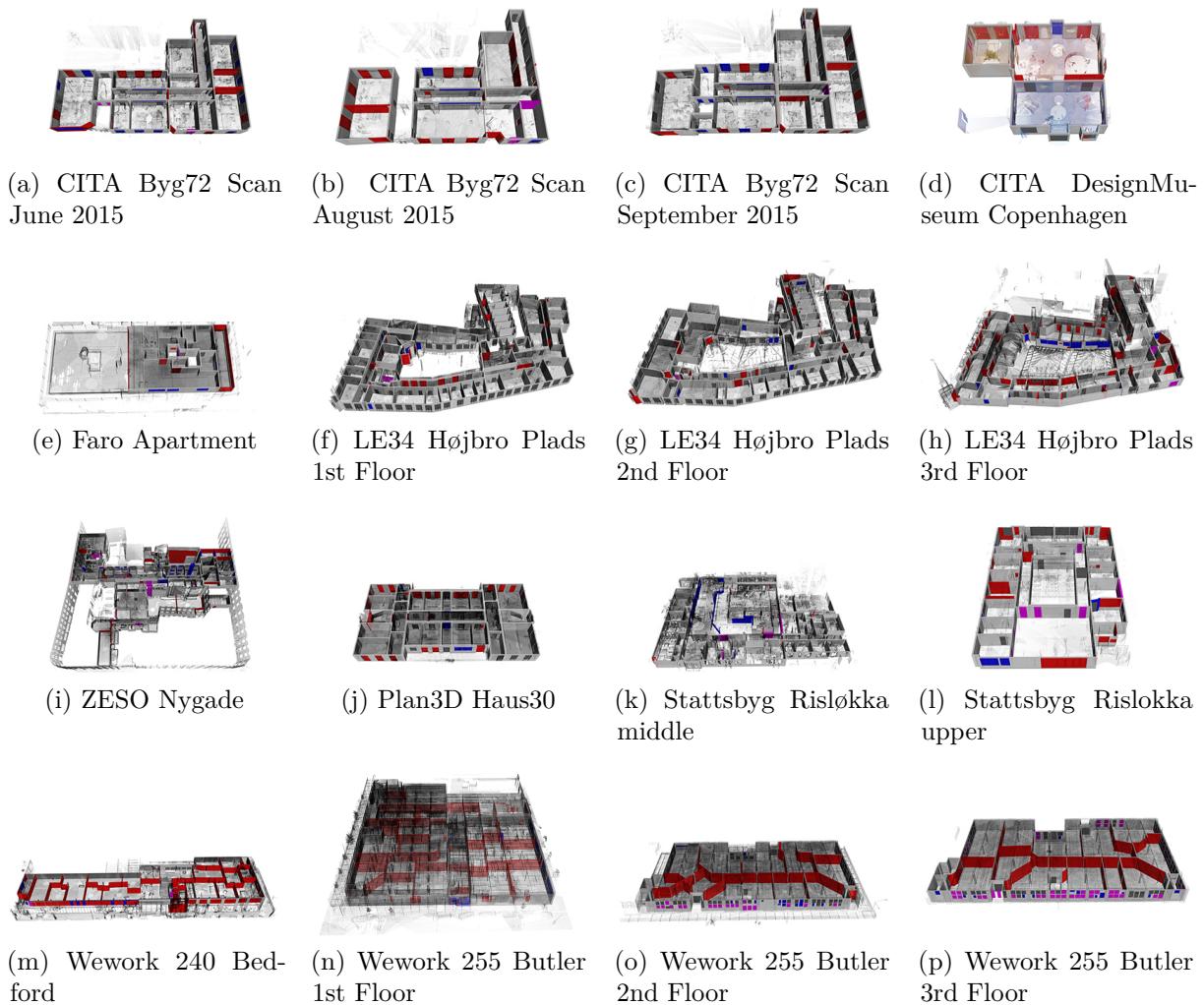


Figure 34: Images of colour coded reconstructed IFC models with Point Cloud overlay.

This data was used to calculate the differences in percentage of False reconstructions and Missing reconstructions to the total amount of elements.

DataSet	False Wall	Missing Wall	False Door	Missing Door	False Window	Missing Window
CITA Byg72 Scan June 2015	5.7	0.0	38.2	2.9	17.2	3.3
CITA Byg72 Scan August 2015	8.0	2.0	58.3	0.0	30.0	4.8
CITA Byg72 Scan September 2015	1.7	1.6	45.9	2.6	14.3	0.0
Plan3D Haus30	1.2	0.0	47.4	0.0	21.1	0.0
Statsbyg Risløkka upper	3.1	0.0	17.6	19.0	28.1	0.0
Faro Apartment	2.5	2.5	31.3	0.0	46.7	0.0
LE34 Højbro Plads 1st Floor	0.9	0.4	30.5	2.1	5.5	1.4
LE34 Højbro Plads 2nd Floor	3.8	0.0	35.7	1.2	11.5	0.0
LE34 Højbro Plads 3rd Floor	8.3	0.6	50.0	2.0	64.2	1.2
CITA DesignMuseum Copenhagen	18.2	0.0	35.7	6.7	40.0	0.0
Statsbyg Risløkka middle	7.0	1.7	44.6	5.1	50.6	3.5
ZESO Nygade	2.4	0.9	38.1	3.0	22.8	4.2
Wework 240 Bedford	37.9	2.2	97.0	0.0	22.4	1.3
Wework 255 Butler 1st Floor	26.7	0.0	74.1	0.0	82.0	0.0
Wework 255 Butler 2nd Floor	34.3	0.0	87.5	0.0	1.7	25.3
Wework 255 Butler 3rd Floor	18.5	0.0	89.8	0.0	0.0	49.6

Table 4: Difference in Percentage of False reconstructions and Missing reconstructions of elements to the total amount of elements reconstructed. A chart is depicted in the Results Section fig. 36,40,50

Statistical Analysis A statistical analysis of the amount and area of reconstructed elements against the 'ground truth' of the models from stakeholders is undertaken. The outcome is the difference in percentage of an element type (wall, door, window) between the reconstructed IFC and the stakeholder modelled IFC.

DataSet	Original			Corrected		
	Amount	Total Area	Avg. Area	Amount	Total Area	Avg. Area
CITA Byg72 Scan June 2015	20.5	-12.0	-27.0	13.6	-14.3	-24.6
CITA Byg72 Scan August 2015	13.6	-19.1	-28.8	4.5	-23.0	-26.4
CITA Byg72 Scan September 2015	36.4	-14.7	-37.4	34.1	-15.2	-36.8
Plan3D Haus30	14.9	-14.6	-25.6	13.5	-15.7	-25.7
Stattsyg Rislokka middle	70.1	6.6	-37.3	59.7	2.9	-35.6
Stattsyg Rislokka upper	52.4	-18.4	-46.5	47.6	-19.7	-45.6
Faro Apartment	2.6	25.5	22.3	2.6	25.5	22.3

Table 5: Differences between the reconstructed IFC and the 'ground truth' IFC in percentage of the walls.

Amount = % of the amount of elements

Total Area = % of the total area of elements

Avg. Area = % of the average area of an element

The 'Original' implies that this is calculated for the original IFC reconstruction and the 'Corrected' implies that the percentages of the False reconstructions have been subtracted.

As chart see fig. 39

DataSet	Original			Corrected		
	Amount	Total Area	Avg. Area	Amount	Total Area	Avg. Area
CITA Byg72 Scan June 2015	61.9	198.3	84.2	0.0	5.2	5.2
CITA Byg72 Scan August 2015	71.4	285.9	125.1	-28.6	-30.2	-2.3
CITA Byg72 Scan September 2015	76.2	280.1	115.7	-4.8	-4.7	0.1
Plan3D Haus30	103.6	236.2	-25.7	7.1	35.9	26.8
Stattsyg Rislokka middle	5.4	155.4	-35.6	-51.4	-38.5	26.3
Stattsyg Rislokka upper	17.2	91.5	-45.6	-3.4	14.7	18.8
Faro Apartment	14.3	97.9	22.3	-21.4	-26.5	-6.4

Table 6: Differences between the reconstructed IFC and the 'ground truth' IFC in percentage of the walls. For further explanation see Table 5. As chart see fig. 44

Windows	Original			Corrected			
	DataSet	Amount	Total Area	Avg. Area	Amount	Total Area	Avg. Area
CITA Byg72 Scan June 2015	-21.6	20.8	54.2		-35.1	-8.2	41.6
CITA Byg72 Scan August 2015	-45.9	-38.1	14.6		-62.2	-54.7	19.8
CITA Byg72 Scan September 2015	-43.2	-15.9	48.2		-51.4	-23.2	57.9
Plan3D Haus30	-48.6	-56.6	26.8		-59.5	-64.4	-12.1
Stattsbyg Risløkka middle	130.0	32.1	26.3		6.7	-49.7	-52.8
Stattsbyg Risløkka upper	77.8	-24.9	18.8		27.8	-39.7	-52.8
Faro Apartment	87.5	322.3	-6.4		0.0	36.5	36.5

Table 7: Differences between the reconstructed IFC and the 'ground truth' IFC in percentage of the walls. For further explanation see Table 5. As chart see fig. 53

Results

Relevance This section describes the relevance of the IFC reconstruction component for the stakeholders.

As mentioned in deliverable D7.7.1, 3D scanning and Point Clouds are commonly used for extensions or renovations of buildings. Here 3D CAD models or BIM models are either modelled from underlying 2D orthographic representations or Point Cloud. The complexity of architecture demands a simplified representation for planning tasks. Here the concept of "as-built" BIM describes a process to obtain a semantically enriched 3D model, representing both physical and functional characteristics of a physical structure [7].

As stated in [7] 'the process of "as-built" BIM is mainly a manual process that can be tedious, intensive and subjective' (fig. 35).

GLOBAL TIME ALLOCATION

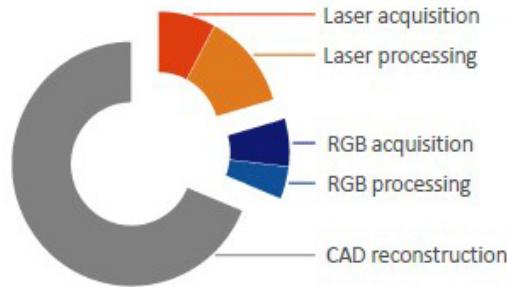


Figure 35: Global time allocation for creating an 'as-built' model from 3D scanning. [8]

An automated process for reconstruction of "as-built" BIM models has the potential to save many work hours. The automated reconstructions for this deliverable have been timed and are as follows:

Dataset	time
CITA Byg72 Scan June 2015	2 min 3 sec
CITA Byg72 Scan August 2015	1 min 31 sec
CITA Byg72 Scan September 2015	2 min 26 sec
Plan3D Haus30	4 min 13 sec
Statsbyg Rislokka upper	1 min 20 sec
LE34 Højbro Plads 1st Floor	23 min 56 sec
LE34 Højbro Plads 2nd Floor	22 min 47 sec
LE34 Højbro Plads 3rd Floor	11 min 15 sec
CITA DesignMuseum Copenhagen	4 min 49 sec
Statsbyg Rislokka middle	7 min 2 sec
ZESO Nygade	6 min 54 sec

The presentation of DURAARK Geometric enrichment components to stakeholders or at events (e.g D8.8 Chapter 4.9 bips BuildingSmart Day) raised always a very large interest. This is obvious looking at the amount of time that can be saved by automation.

The relevance can also be seen in the investments done by the commercial industry to be the leader in the field of Point Cloud to BIM automation. FARO (one of the leading 3D scanner producers on the market) has for example bought Kubit VirtuSurv⁴⁴. This software links Point Clouds to many different BIM environments. BIM elements are

⁴⁴http://us.faro3dsoftware.com/CAD/Products/VirtuSurv/3D_Laserscanning.php

semi-automated modelled through selection of predefined element types and selection of points in the planar image produced by the 3D scanner. Or for example IMAGINiT - Scan to BIM⁴⁵ and ClearEdge3D⁴⁶ have developed plugins for Autodesk Revit, which deliver user supported automatic reconstruction of walls.

Precision This section describes the results of the manual comparison and the statistical analysis in terms of reconstructions and labellings of IC elements, **walls**, **doors** and **windows**, and it's dimensions.

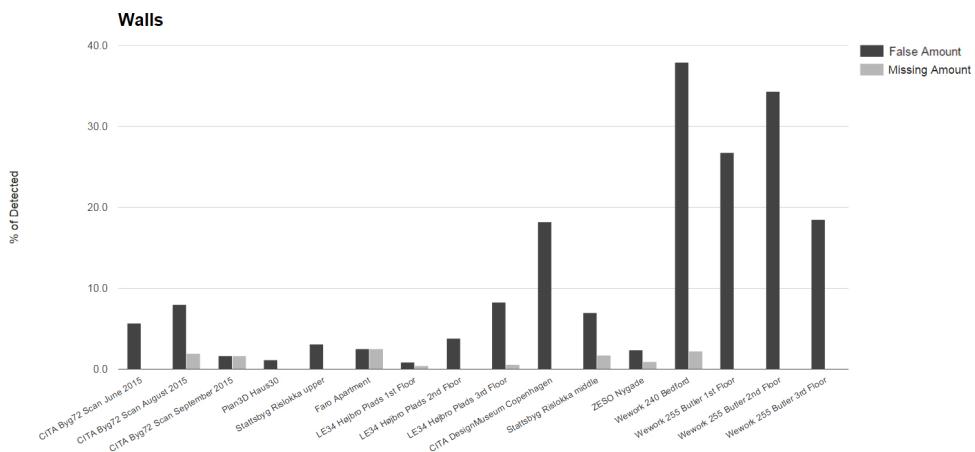


Figure 36: The percentage of reconstructed walls labelled as False or Missing.

The manual comparison between **wall** reconstructions and the Point Cloud shows, that approximately all walls are reconstructed. The results depicted in fig. 36 shows that the amount of Missing walls is in average 0.8% and overall less than 2.5% for all reconstructed IFCs.

The amount of False Wall Reconstructions is as well relatively low. Figure 36 shows, that the average amount of False reconstructions is 11.3%, but for most projects the percentage is below 8%. It can be determined that the main spikes in the chart result from sets of columns and beams being reconstructed as walls (Wework 240 Bedford, Wework 255 Butler 1st-3rd Floor) (see fig. 37) or very thick walls being reconstructed as multiple wall elements (CITA DesignMuseum Copenhagen) (see fig. 37). If the average amount is calculated without taking these projects into account, the amount of False Wall

⁴⁵<http://www.clearedge3d.com/products/edgewise-building/>

⁴⁶<http://www.clearedge3d.com/products/edgewise-building/>

Reconstructions is only 3.7%.

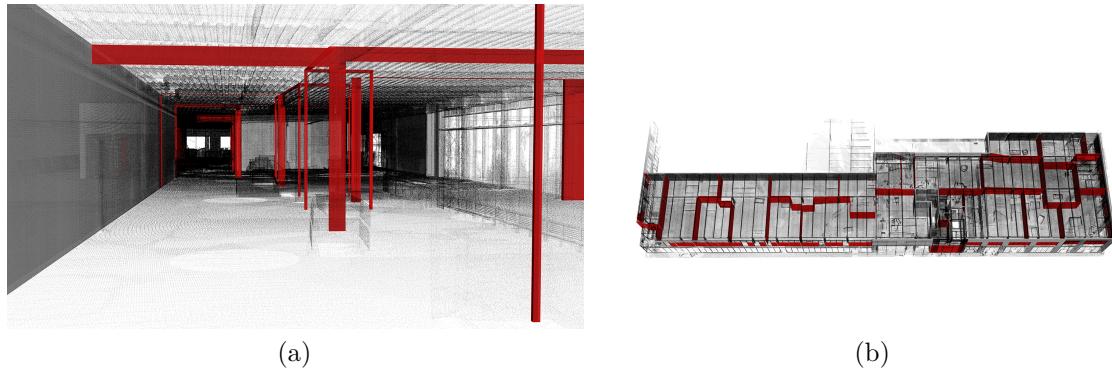


Figure 37: (a) Interior view of False reconstructions of column and beam as walls. (b) Overall view of Wework 240 Bedford with false column and beam reconstructions as walls.

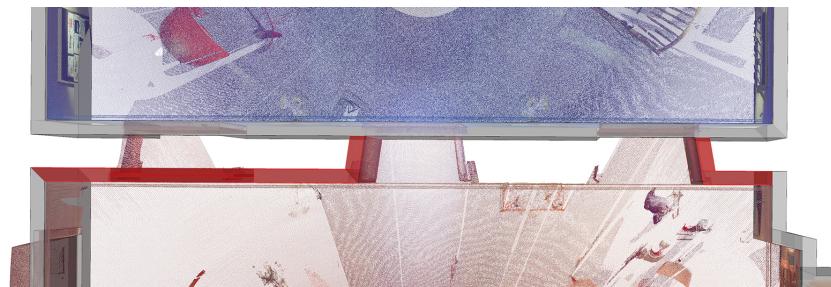


Figure 38: Thick walls reconstructed as multiple wall elements in CITA DesignMuseum Copenhagen.

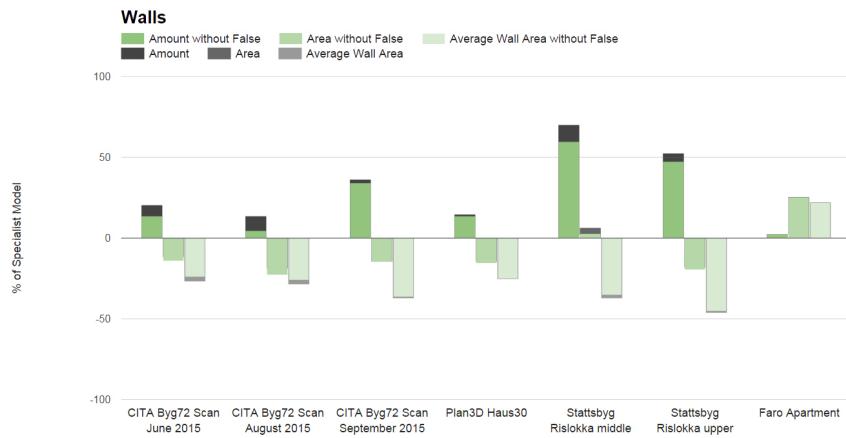


Figure 39: The statistical analysis of differences in the reconstructed wall elements and it's 'ground truth'. Gray displays the original deviation between the reconstruction and the 'ground truth' in percentage. Green displays the deviation without taking False reconstructions into account. The large deviation between Statssbyg Rislokka and the 'ground truth' is due to cubicals, not drawn in the 'ground truth', being reconstructed as walls

The statistical analysis of differences in the reconstructed wall elements and it's 'ground truth' shows, that in average 30.0% more wall elements are reconstructed. The total area of Wall Reconstructions is in average 6.7% smaller than the 'ground truth', and the average wall element size is 25.8% smaller (fig. 39). The total area indicates that the reconstruction of walls is in general under-dimensioned. This is most likely due to thinner exterior walls, because the reconstruction does not take exterior scans into account. The larger amount of walls and smaller average size of wall elements indicates that the reconstructed walls are segmented into smaller pieces than what is present in the 'ground truth'.

The green part of the columns in fig. 39 indicates the percentage of walls deviating from the 'ground truth' after subtracting wrongly reconstructed wall elements.

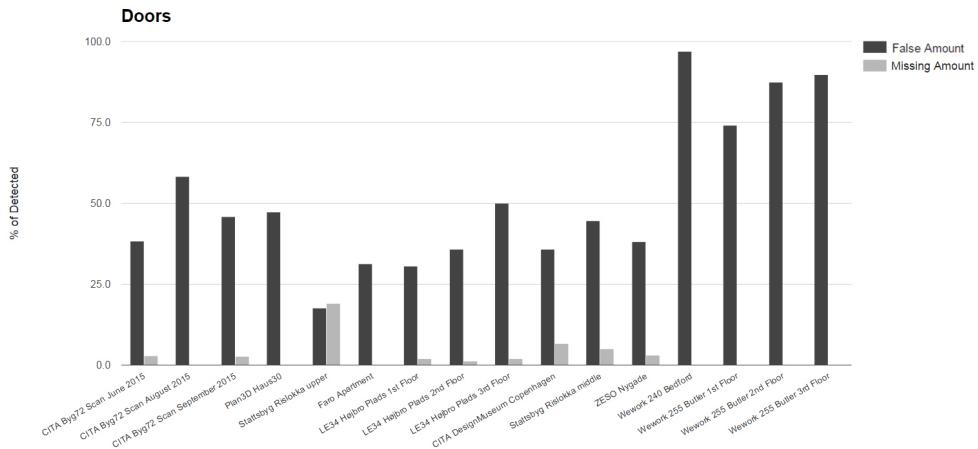


Figure 40: The percentage of reconstructed doors labelled as False or Missing.

Figure 40 shows that almost all **doors** are reconstructed. The average amount of Missing Door Reconstructions is 2.8%. This amount does not take the False Positive window reconstructions into account. This distinction has been made because these doors actually have been detected as openings, but have been labelled falsely as windows. These False labellings appear in the statistics of the windows.

The main reason for the missing reconstruction of doors can be traced to poor scanning practice, where doors have been closed at the time of scanning (e.g. Statsbyg Risløkka upper) (fig. 41). If this single bad dataset (Statsbyg Risløkka upper) is not taken into account the average amount of Missing Door Reconstructions falls to only 1.7%.

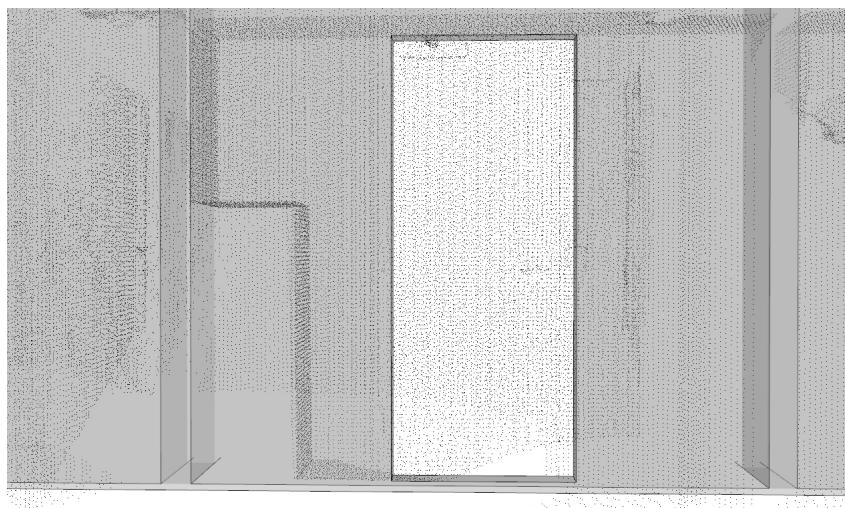


Figure 41: Example of Missing Reconstruction of door because of a closed door, while scanning. Statsbyg Risløkka upper.

The comparison showed a generally large percentage of False Door Reconstructions with an average of 51.4%. These False reconstructions are mainly due to False Positive Door Reconstructions, where a window opening has been detected but labelled as a door (e.g. Plan3D Haus30) (fig. 42).

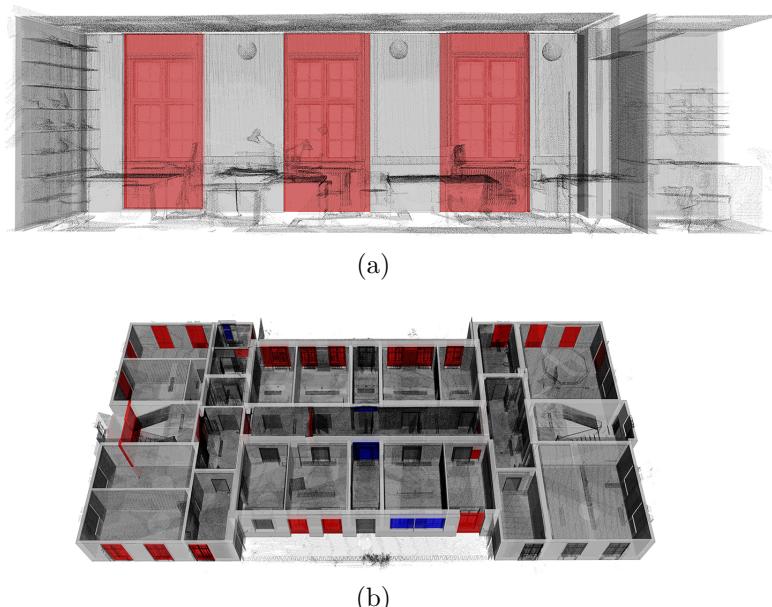


Figure 42: (a) Interior view of False Positive reconstructions of doors, where a window opening has been detected but labelled as a door in Plan3D Haus30. (b) Overview of Plan3D Haus30 with annotations. Majority of Red annotations in exterior walls are False Positive door reconstructions.

Another reason for False Door Reconstructions stems from the previous described False Wall reconstruction of sets of columns and beams, where a large door is reconstructed in the void between the columns and beams (e.g. Wework 240) (fig. 43).

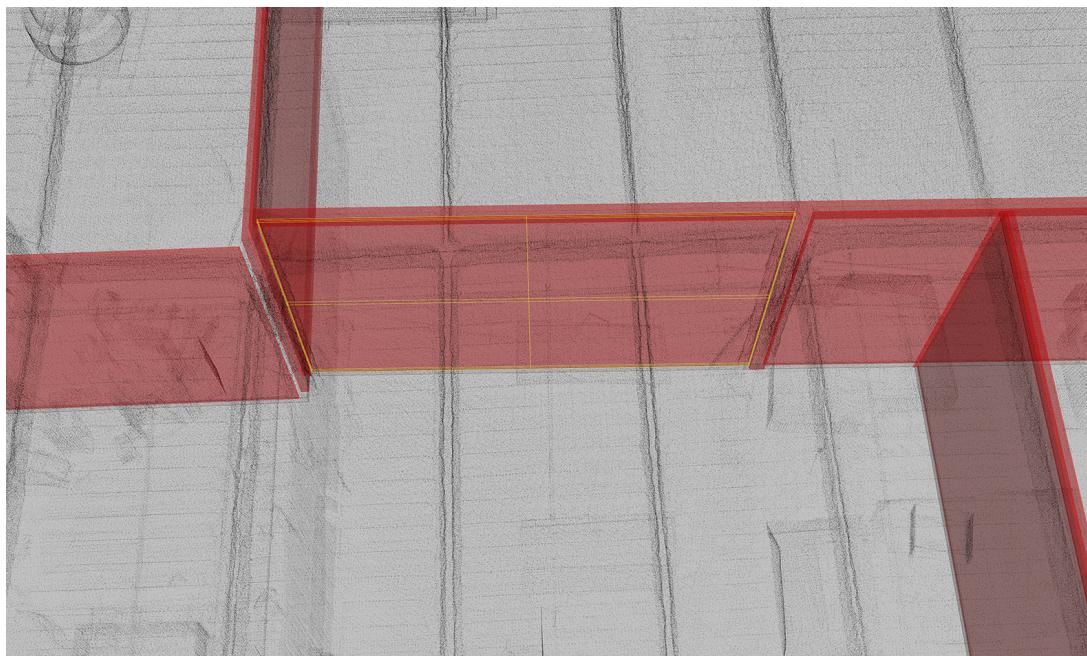


Figure 43: False Door Reconstructions of sets of columns and beams, where a large door is reconstructed in the void between the columns and beams. Wework 240.

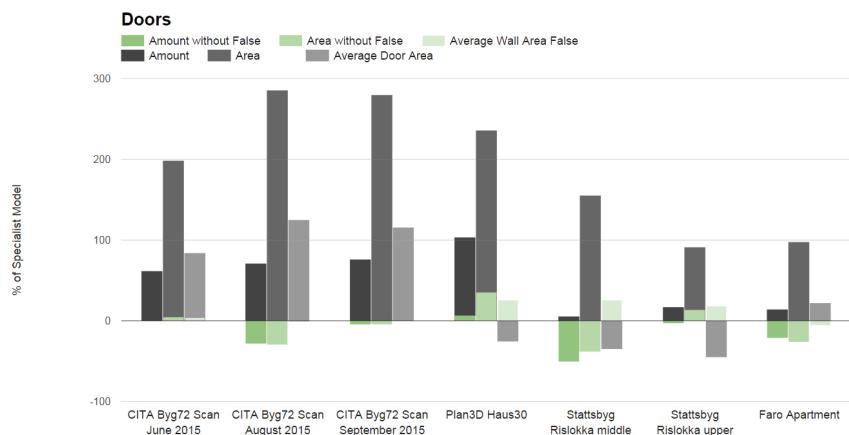


Figure 44: The statistical analysis of differences in the reconstructed door elements and it's 'ground truth'. Gray displays the original deviation between the reconstruction and the 'ground truth' in percentage. Green displays the deviation without taking False reconstructions into account.

The analysis of differences between reconstructed doors and the 'ground truth' shows, that the average difference in the amount of elements is 50.0%, the total area is 192.2% and the average element area is 34.4% larger. As previously mentioned this is mainly

due to False Positive Door Reconstructions. When these False Door Reconstructions are not taken into account the picture is very different. Then the average amount of doors is 14.6% smaller, the total area is 6.3% smaller and the average element area is 9.8% larger (fig. 44). This corresponds with the low amount of Missing Door Reconstructions described earlier and indicates a large False Positive Door Reconstruction and False Positive Window Reconstructions. It also shows that the correctly reconstructed doors in general are slightly over dimensioned (fig. 45).

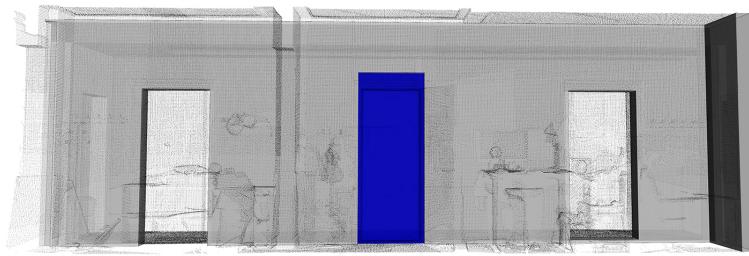


Figure 45: View of general example of eversized deviating reconstructed door. The reconstructed door deviates in height.

Figure 46 and Figure 47 show histograms of the width and height of all reconstructed doors compared with 33 IFC 'real world' IFC models from the DURAARK repository. This gives a clear picture that the width and height of reconstructed doors are in general over dimensioned. When zooming in on the histograms (fig: 48 and 49) it is possible to see a pattern of plateaus in the 'green' histogram of the repository data. This indicates that the most common used doors have a width of 700mm, 900mm or 1000mm and almost all doors have a height of 2100mm. Data like this indicates that a heuristic approach to eliminate False door candidates could be integrated into the reconstruction algorithm.

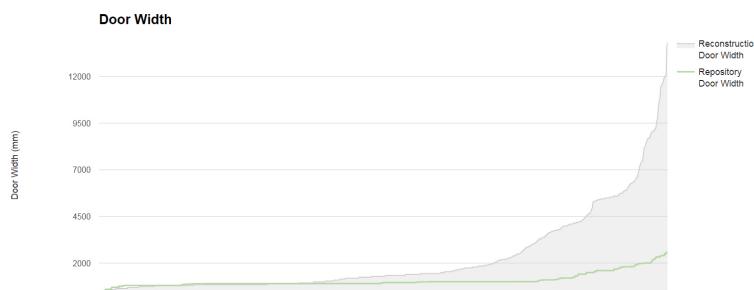


Figure 46: Reconstructed door widths compared with door widths of 33 'real world' IFC from DURAARK repository.

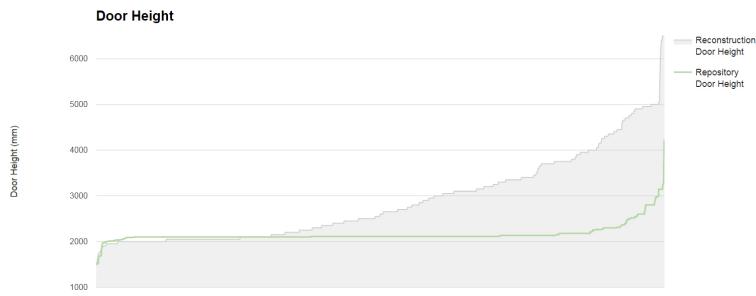


Figure 47: Reconstructed door heights compared with door heights of 33 'real world' IFC from DURAARK repository.

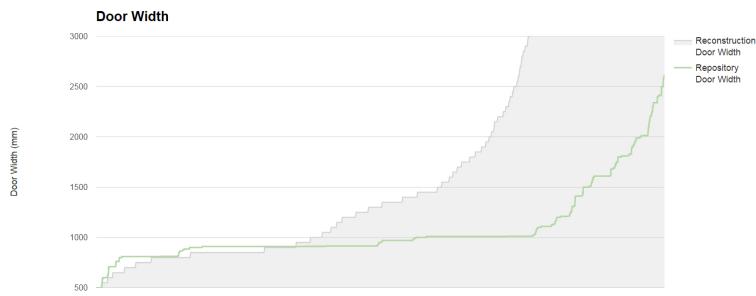


Figure 48: Reconstructed door widths compared with door widths of 33 'real world' IFC from DURAARK repository. Zoomed in.

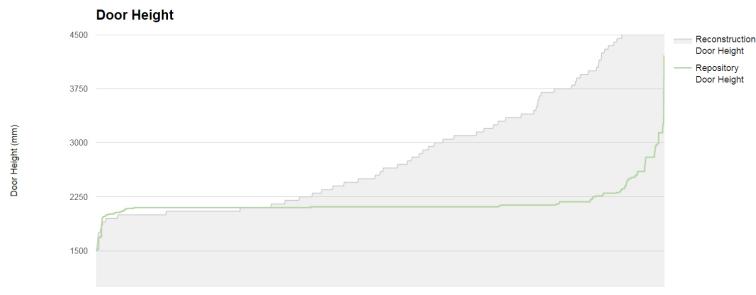


Figure 49: Reconstructed door heights compared with door heights of 33 'real world' IFC from DURAARK repository. Zoomed in.

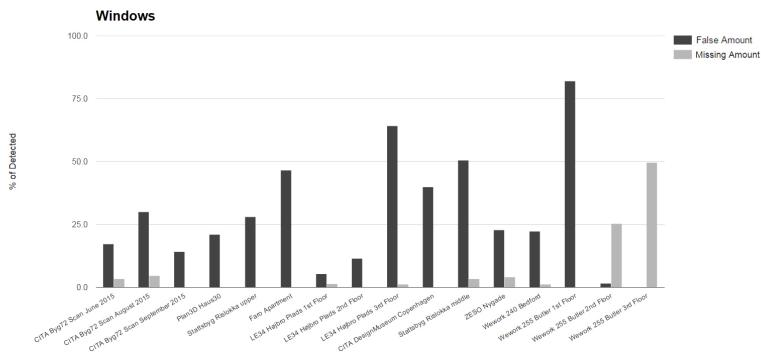


Figure 50: The percentage of reconstructed windows labelled as False or Missing.

The same pattern emerges when we investigating the **windows**. Figure 50 shows a relatively small amount of Missing Window Reconstructions with an average of 5.9%. This amount does not take the False Positive door reconstructions into account. These False labellings appear in the statistics of the doors.

There are mainly two reconstructions, which contain Missing Window Reconstructions. These reconstructions are found in WeWork 255 Butler 2nd-3rd Floor, which is an old factory building (fig. 51)

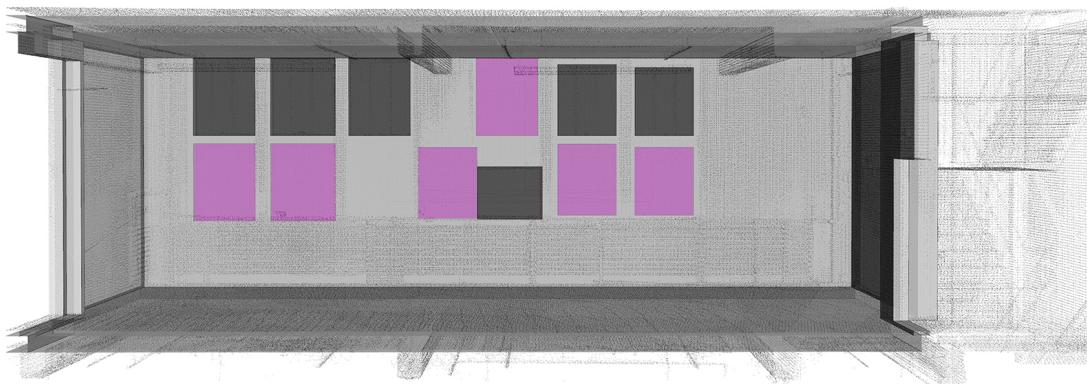


Figure 51: View of Missing reconstructions of windows in a wall of Wework 255 Butler

Generally there a large percentage of False Window Reconstructions can be observed. This is due to False Positive Window Reconstructions, where a door opening has been detected but labelled as a window (e.g. fig. 52).

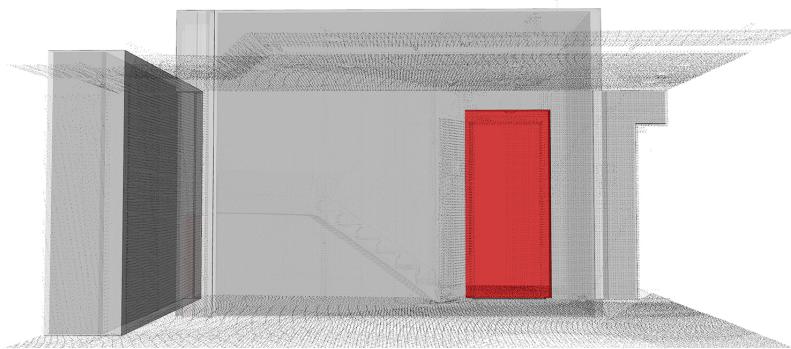


Figure 52: View of False Positive Window Reconstructions, where a door opening has been detected but labelled as a window.

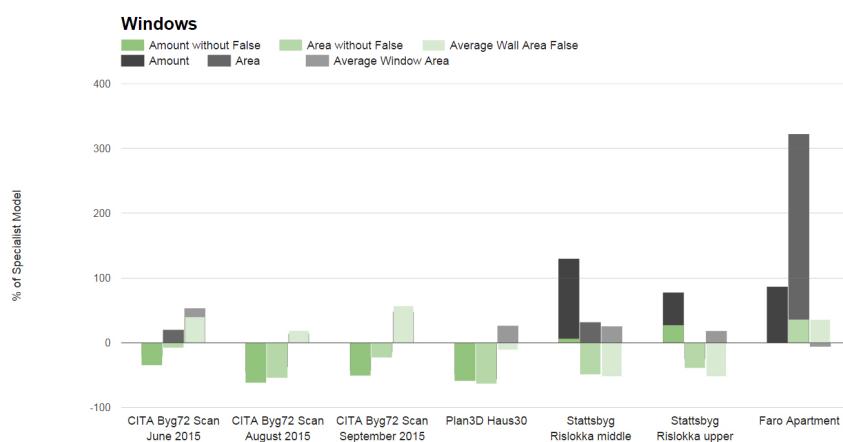


Figure 53: The statistical analysis of differences in the reconstructed window elements and it's 'ground truth'. Gray displays the original deviation between the reconstruction and the 'ground truth' in percentage. Green displays the deviation without taking False reconstructions into account.

The analysis of differences between reconstructed windows and their 'ground truth' shows, that the average difference in the amount of elements is 19.4% larger, the total is 34.3% larger and the element area is 26.0% larger. If the False Reconstructions are not taken into account, the difference in the amount is 24.8% smaller, the total area is 29.0% smaller and the element area is 5.4% larger.

This corresponds to the low amount of Missing Windows Reconstructions and the previously mentioned large amount of False Positive Door Reconstructions. When investigating the original statistic and the corrected ones the large amount of False Positive Window Reconstructions becomes apparent. It also shows that the average size of Window Reconstructions is in general over dimensioned (fig. 54).

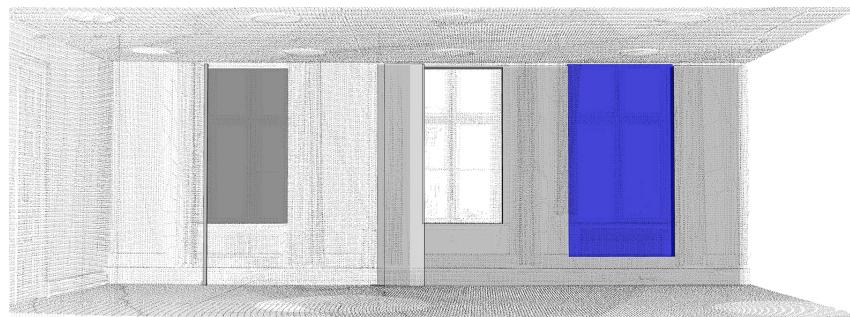


Figure 54: View of general example of eversized deviating reconstructed window. The reconstructed door deviates in height.

Figure 55 and figure 56 show a histogram of the widths and heights of all reconstructed windows compared with the widths and heights of windows of 33 'real world' IFC models from the DURAARK repository. The widths and height of the reconstructed windows follows very well the histogram of the 'real world' data. It is though possible to distinguishes plateaus at about 1200mm, 2300mm and 2700mm for widths and one at 1200mm for heights of the 'real world' data. This indicates that data like this could be used in a heuristic approach to eliminate False window candidates. It is though not as pronounced as for doors.

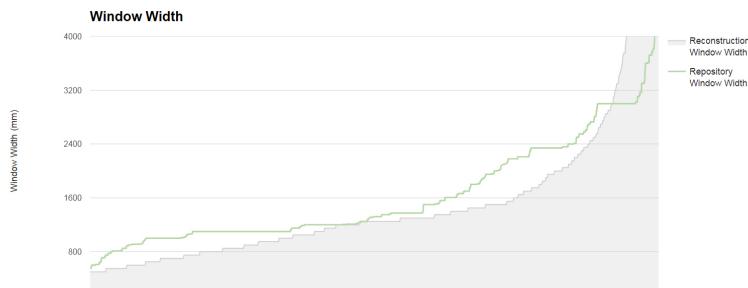


Figure 55: Reconstructed window widths compared with window widths of 33 'real world' IFC from DURAARK repository.

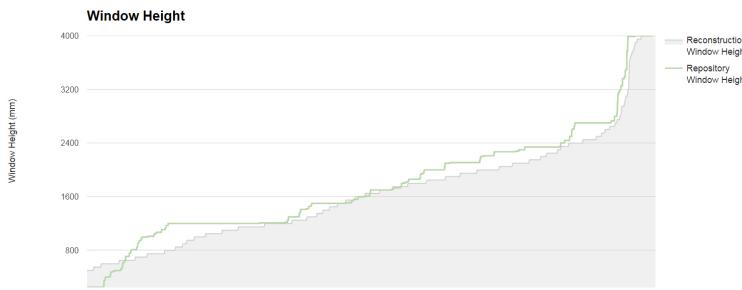


Figure 56: Reconstructed window heights compared with window heights of 33 'real world' IFC from DURAARK repository.

Quality The overall **completeness** and **tolerance** of the reconstructed IFC's is very close to the requirements of the stakeholders. The main failures to comply with these requirements are the amount of False Positive Reconstructions, where the main reason is the labelling of windows as doors. The amount of Missing Reconstructions is relatively small and is not seen as a major factor for meeting the stakeholders' requirements.

Concerning **drawing conventions**, which were criticised in D7.2, several improvements are observed. These improvements have been made by eliminating the short wall segments by means of a "short wall" mitigation in the IFC reconstruction Component. The reconstructions display still a higher amount and smaller area of wall segments than in the 'ground truth' by stakeholder. This indicates that there is still a difference in the abstraction level and drawing conventions, but with a clear improvement from D7.2. A similar mitigation approach could be conducted on floor and roof slabs of the reconstructions. Comparing the reconstructions with the 'ground truth' shows that the reconstructions has a slab for each room, while the 'ground truth' in general has a slab per level (fig. 57).

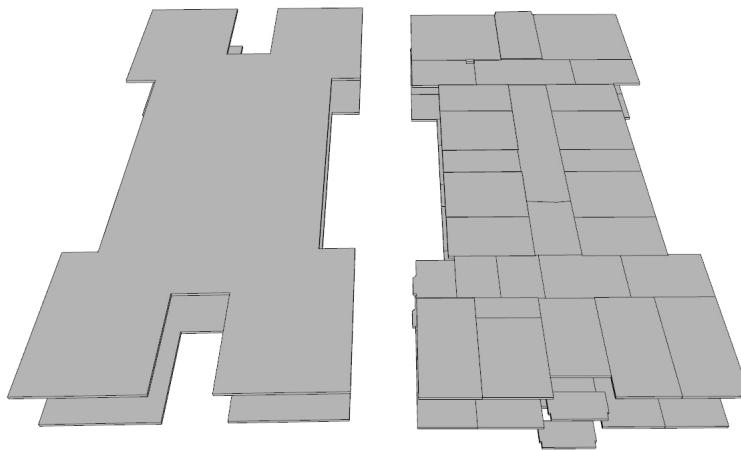


Figure 57: Slabs of Plan3D Haus30. Left depicts the slabs of the 'ground truth'. Right depicts the slabs of the reconstructions.

7.3 DURAARK components and Novel 3D Acquisition Techniques

The field of 3d registration techniques is constantly expanding. In this chapter we are revisiting the decision of the DURAARK consortium to focus on 3d Laserscanning as acquisition technique for Point Clouds as described in D2.1. We investigate in how far datasets, which originate from alternative acquisition techniques, can be processed with the developed DURAARK IFC reconstruction components.

We base this investigation on two techniques:

- **LIDAR** - a 3D acquisition commonly used for aerial surveillance, was lately introduced as a indoor scanning technique in several commercial products ⁵⁸. These promises to be as accurate, but way faster than 3d Laserscanning. The technique is used in products like the ZEB1 Hand-held Laser Scanner⁴⁷ or the Leica Pegasus:Backpack⁴⁸. Current advances in technological developments among industry ⁴⁹⁵⁰ suggest, that LIDAR Scanners will become way cheaper and more accurate in the near future.

⁴⁷<http://www.3dlasermapping.com/zeb1/>

⁴⁸<http://leica-geosystems.com/products/mobile-sensor-platforms/capture-platforms/leica-pegasus-backpack>

⁴⁹<http://www.velodynelidar.com/>

⁵⁰<http://www.quanergy.com/>



Figure 58: The ZEB1 Hand-held 3d sensor and the Leica Pegasus Backpack are commercial indoor 3d registration techniques, that are mobile and faster than 3d laserscanning.

- **Structure From Motion (SFM)** - This approach uses images of objects and spaces as input for the creation of Point Clouds. Advances in the computation of images made this approach suitable for application on mobile devices and consumer camera products. The approach has been popularised for the creation of 3d models from drone based aerial photography⁵¹. Current research suggests furthermore an increase of the resolution of camera based 3d acquisition techniques⁵². Camera based approaches are cheap and several research prototypes, as VisualSfM⁵³, exists besides solutions, that utilize the Kinect sensor^{54 55 56}. As the range of the sensor is limited (not more than 3meters) it is mainly used for the capturing of objects. Several commercial products exist for the 3d registration of buildings and spaces , such as Bentley ContextCapture⁵⁷, Autodesk Recap⁵⁸ and Agisoft⁵⁹.

Based on real world datasets and the experience of parsing these through the DURAARK components, we discuss in this chapter finally how suitable the above mentioned approaches are for the creation of Point Clouds of indoor spaces and how both, the DURAARK

⁵¹<https://pix4d.com/>

⁵²<http://news.mit.edu/2015/algorithms-boost-3-d-imaging-resolution-1000-times-1201>

⁵³<http://ccwu.me/vsfm/>

⁵⁴<http://reconstructme.net/>

⁵⁵<http://www.kscan3d.com/>

⁵⁶<http://3dprint.com/20780/microsoft-kinect-3d-builder/>

⁵⁷<https://www.bentley.com/en/products/brands/contextcapture>

⁵⁸<http://www.autodesk.com/products/recap/features/photo-to-3d/gallery-view>

⁵⁹<http://www.agisoft.com/>

tools and the acquisition techniques, need to improve in order to generate meaningful architectural data from the in this way acquired Point Clouds.

Aspects evaluated

Utility - Are the DURAARK tools able to include alternative 3d aquistion technologies?

Utility - What is the effort to create datasets from alternative acquisition techniques, that are goof enough as input to the developed techniques? How can the DURAARK geometric enrichment tools be adopted to the alternative acquisition technologies? What are the key necessary input parameters for the geometric enrcihment tools and how can these be extracted from the alternative acquisition technologies?

Method

Tests with datasets from collaboration with Construction Company - Monitoring over time (Point Cloud through photgrammteric appraoches)

Datasets and Evaluation Activities

The following datasets have been used for the evaluation activities:

The dataset COWI-SLAM 1 & 2 were created with a **SLAM based** ZEB1 handheld scanner from the large exhibition hall at KADK⁵⁹. Scanned Points in the SLAM dataset do only consist of XYZ coordinates. No other parameters, such as intensity, normal or the position they were registered from, are stored with them. Three indoor datasets from **Structure from motion** were acquired:

- Nyborg Strand⁶⁰
- KADK Festsalen
- KADK Bygning 72

All three have been created by experts in photogrammetric technology and were processed with the SFM software agisoft into e57 PointClouds. The three datasets are all only showing fragments of the spaces and the Point Cloud representations of architectural surfaces are very noisy. This despite the fact, that for the SLM process big efforts were undertaken to create well lit, focused images with high contrasts. In order to gain this

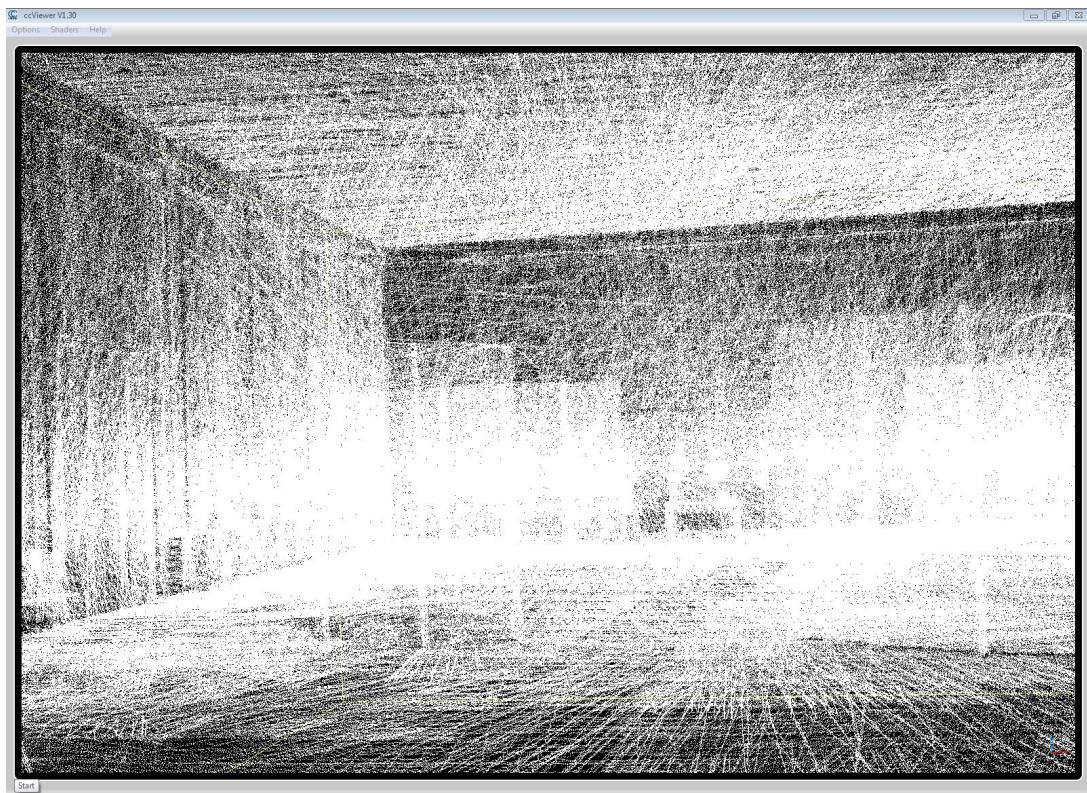


Figure 59: SLAM technology was used to create the dataset COWI-SLAM 2- a scan of the KADK exhibition hall. The data shows the lines of scanned points - characteristic for the ZEB1 handheld scanner.



Figure 60: The interior of the Nyborg Strand hotel was registered using a SFM software. The approach has big problems to create interior Point Clouds from interior images, which are characterised by low light, high contrast range and lack of traceable patterns in the different images.

CITA used a set of photographic technologies, that should ensure high quality (use of tripod, HDRI bracketing, Flash, indirect light). Which factors were the main contributors to the bad quality of the created Point Clouds, could not finally be resolved:

- Lack of contrast on surfaces
- Uniform wall surfaces do not provide enough pattern for SFM image algorithms
- Too high contrast in images between areas with light and shadow (over- and underexposure)
- Too many dark, underexposed areas with little contrast
- Too little camera positions in too small rooms

The outcome of the SFM approach are datasets, that challenge the IFC reconstruction tool through their noisiness, incompleteness and lack of basic parameters, that the approach is based on.

Results

COWI-SLAM 1 The reconstruction results of the COWI-SLAM 1 dataset are shown in fig. 61. The dataset consists of 5 rooms which are available as separate, registered scans in one E57 file. Although the rooms are stored as separate scans, it should be noted that the shapes of the scans do not fulfill certain assumptions which can be made when dealing with scans from stationary laser scanners (with which the reconstruction prototype was originally tested during development and evaluation): In particular, non-convex rooms (e.g. the blue scan in the figure) or multiple, vertically stacked rooms (e.g. the magenta scan in the figure) could not be scanned from a fixed position using a conventional laser scanner. Especially the case of multiple stacked rooms which have a single scan label violates the assumptions made in the current IFC reconstruction approach. Additionally, the dataset contains rooms which are overlapping when projected down to the horizontal plane (e.g. the blue and yellow rooms in the figure) which cannot currently be handled correctly. Nonetheless, the reconstruction prototype performs a reconstruction which represents the room geometry quite well, except for regions where e.g. the blue and yellow scans are stacked. Some floor and ceiling heights are not estimated correctly, possibly due to a too sparse sampling of the geometry in the Point Cloud.

COWI-SLAM 2 Fig. 62 shows the reconstruction results of the COWI-SLAM 2 dataset. The whole dataset is stored as a single scan although multiple rooms were measured. Since the reconstruction approach assumes at least one scan per room in order to perform a labeling of separate rooms, the whole building is inevitably reconstructed as a single room. The overall shape – especially of the large room – was captured quite well. On the right hand side of the image, a large spike in the reconstructed wall is visible which can be explained by missing wall geometry in the Point Cloud data at this location. This is especially visible in the bottom right part of the figure: The scan “fades out” without being clearly bounded by a wall. In this case, the reconstruction algorithm tries to close this hole in some way but does not have a good guidance by measured wall geometry.

Nyborg Strand The Nyborg Strand dataset is a scan of a single, relatively simple room, the reconstruction results are shown in fig. 63. Given the relatively high noise level of the Point Cloud, the reconstructed walls are positioned fairly well, without being overly segmented. One particular problem of this dataset is that the floor has very large holes, and the ceiling structure is not contained in the Point Cloud measurements at all. As

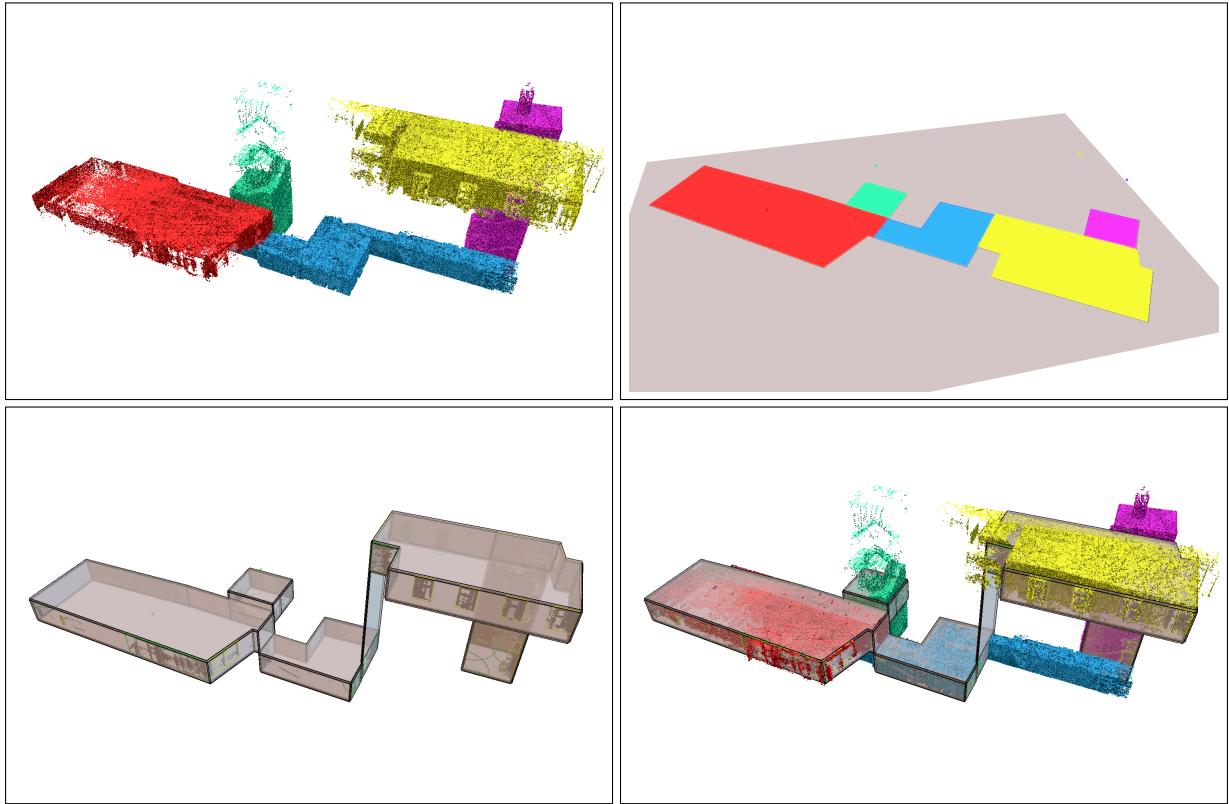


Figure 61: Reconstruction results of the COWI-SLAM 1 dataset. Left to right, top to bottom: Input Point Cloud with 5 separate scans, reconstructed room layout, reconstructed model, hybrid visualization of the model and Point Cloud.

a consequence, the floor and ceiling heights cannot be estimated by the reconstruction algorithm which results in an incorrect vertical placement of the room as can be seen in the bottom-right part of the figure. However, this is a problem which can quickly and easily be fixed manually.

KADK Festalen The reconstruction results of the KADK Festalen dataset are shown in fig. 64. The main problem of this scan of a single, relatively big room is that large parts of the walls, but also parts of the floor and ceiling are missing as can be seen in the upper-left part of the figure. This results in problems when performing the reconstruction: The approach assumes that each wall which should be considered at least as a candidate for a wall in the final model needs to be measured at least partially. In case of the dataset, the side of the room shown in the upper-left corner in the figure is not bounded by any measured wall geometry. This is especially visible in the right hand part of the

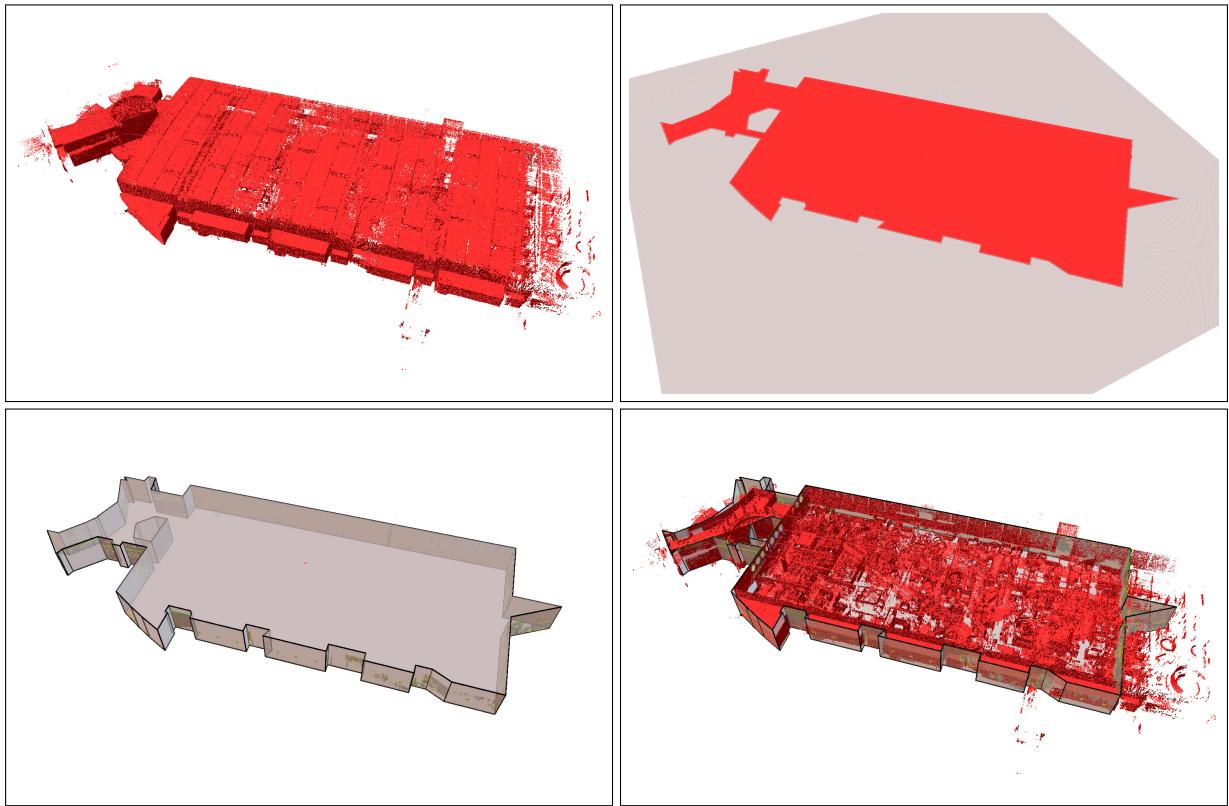


Figure 62: Reconstruction results of the COWI-SLAM 2 dataset. Left to right, top to bottom: Input Point Cloud, reconstructed room layout, reconstructed model, hybrid visualization of the model and Point Cloud.

image where the ceiling structure has been hidden. As a consequence, the geometry of the reconstructed room “bleeds out” and is only restricted by the bounding box of the area in which the room layout is determined. Since this is not expected behavior, the reconstruction does not correctly reconstruct wall geometry. Thus only the room layout is shown in the figure.

Conclusion

The IFC reconstruction prototype developed by UBO has been evaluated on several non-trivial Point Cloud datasets captured using SLAM and structure from motion approaches. The achieved results are promising given the facts that (1) the prototype was originally developed, tested and evaluated on high-quality, high-resolution Point Cloud scans from stationary laser scanners, and (2) the datasets evaluated here clearly violate some assumptions which were valid for datasets for which the software was developed. In

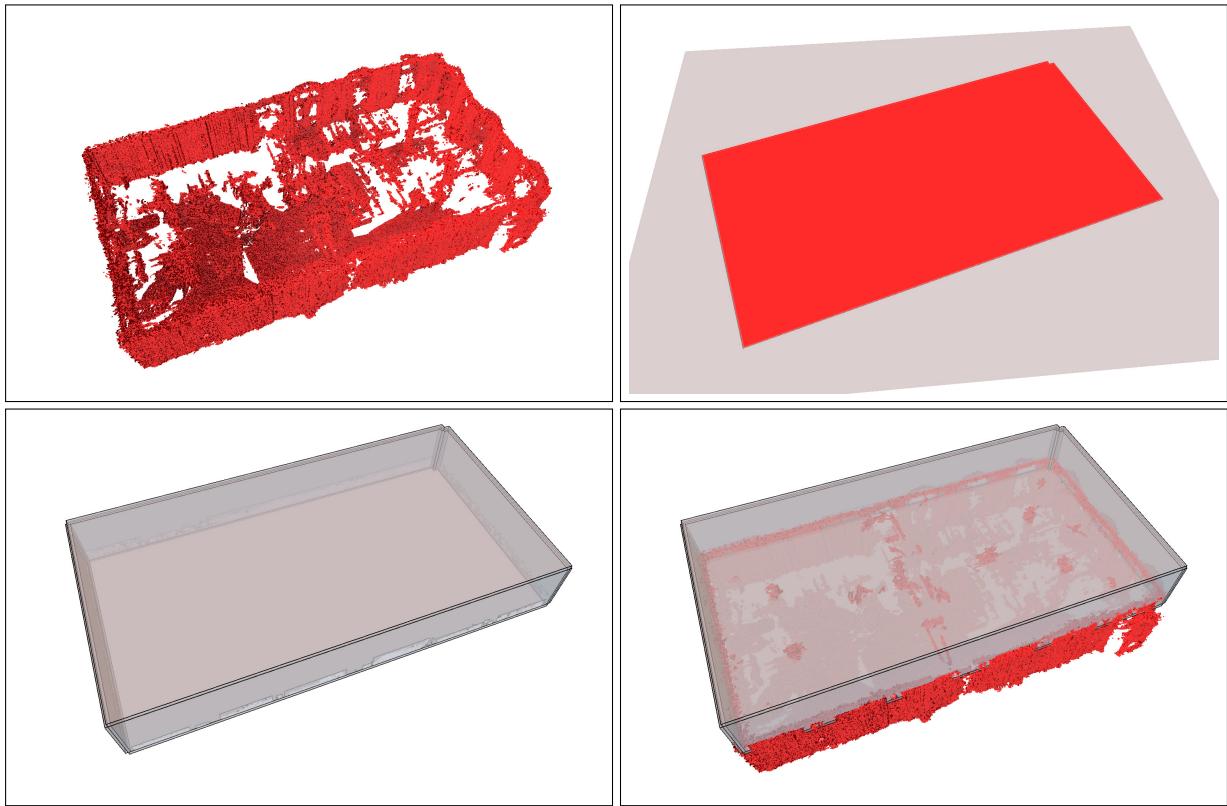


Figure 63: Reconstruction results of the Nyborg Strand dataset. Left to right, top to bottom: Input Point Cloud, reconstructed room layout, reconstructed model, hybrid visualization of the model and Point Cloud.

addition, the quality of the scans, in particular regarding missing data and noise levels, would at least require some parameter tuning which has not been done for this evaluation, i.e. the same parameters as for stationary scanning devices were used. Suitable parameters as well as (semi-automatic) means to guide the algorithm where e.g. walls are completely missing would most probably improve the results.

The tested acquisition approaches promise faster and in the case of SFM cheaper 3d registration. While SFM might be cheap in terms of equipment and software needed and is well suited for outdoor photogrammetry it failed in its current state to create acceptable results for indoor situations. Despite high technical efforts and a time consumption, that was way higher for the actual capturing, as the post-processing of the data, than that needed for a 3d Laserscanning, results were noisy and incomplete.

The SLAM based approach showed way better results and was extremely fast in comparison to 3d laserscanning. For a good integration with the developed IFC reconstruction

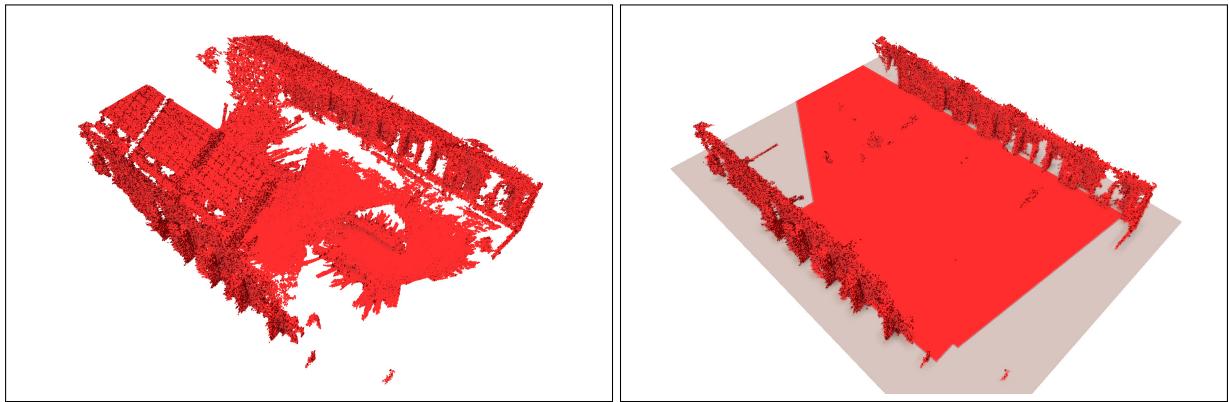


Figure 64: Reconstruction results of the KADK Festsalen dataset. Left: Input Point Cloud. Right: Hybrid visualization of the Point Cloud (ceiling has been hidden manually) and determined room layout.

technique, the points in the datasets lack however information of the position they were scanned from. These parameters are taken during the registration process, but not exported into the final e57 dataset. Doing so bears chances to create a good and fast alternative to the currently used 3d laserscanning acquisition for the DURAARK datasets.

7.4 RISE

This part contains the evaluation aspects of the RISE (Reveal Invisible StructurEs) component. This component consists of several sub-components that will be evaluated separately

- data acquisition and preprocessing
- electrical appliance detection (ElecDetect)
- electrical wiring hypothesis (WireGen)

for a thorough description of the RISE pipeline we refer to deliverable D5.6 .

Aspects evaluated

Data acquisition and preprocessing.

The input of RISE consists of an floorplan description, e.g. provided by the IFC reconstruction service, as well as spherical panoramic images acquired at the pointcloud

Name	# rooms	# scans	walls	groundtruth	wiring
Nygade1001	1	1	6		yes
Nygade1002	1	1	7		no
Nygade1005-1006	1	2	16		no
Nygade37	1	1	6		no
LysLab03	1	1	4		no
Bygade72-2nd	13	14	75		yes

Table 8: The datasets used for evaluation. The number of walls corresponds to the number detected by the IFC extraction.

scanning positions. Evaluation aspects cover quality criterions regarding the input data and preparation effort.

Electrical appliance detection in orthophotos

The used method for image classification were evaluated regarding true positives (correctly labelled), false negatives (missed detections), and false positives (hallucinated elements). Furthermore, the evaluation time was measured.

Electrical wiring hypothesis

precision and completeness of this method depend on the concrete grammar that was used for the hypothesis. The hypothesis of two ground truth datasets was verified.

Method, Datasets and Evaluation Activities

The RISE component was evaluated on the following datasets (see Table 8).

Data acquisition and preprocessing. Effort: The FARO FOCUS 3D scanning unit used to acquire the test data also acquires a panoramic image that can be used to color the point cloud. However, these images typically do not yield high enough resolution for electrical appliance detection, and the sensor is very sensitive to lighting conditions. Therefore, a panoramic high dynamic range image (HDRI) was acquired at the scanners position to account for these shortcomings. For this a F650 Canon Camera with Fish-Eye 8mm f3.5 and inbuilt HDRI bracketing was mounted at the exact position of the 3d laser scanner using a Nodal Ninja 4⁶⁰. See also Figure 66.

⁶⁰<http://shop.nodalninja.com/nodal-ninja-4-1/>

Name	point cloud	floorplan
Nygade1001		
Nygade1002		
Nygade1005-1006		
Nygade37		
LysLab03		
Bygade72-2nd		

Figure 65: Overview of the input point clouds and the extracted floorplan representation.



Figure 66: The FARO FOCUS 3D scanning unit used for acquisition does not yield enough resolution to capture important features that are necessary for classification. Furthermore, the images are easily over- or underexposed when lighting conditions are not optimal (left). Therefore, a manually taken image had to be used that yields higher resolution (right).

Electrical appliance detection in orthophotos For training, a self-annotated database of 41 power sockets and 13 light switches was used. From each annotation, 10 training patches were generated by applying a random jitter, yielding a total number of 540 positive training samples. In order to create negative training data, 726 image patches were randomly extracted from unmarked regions of the annotated set.

Electrical wiring hypothesis

The results of the automatic hypothesis are visually compared to a groundtruth dataset that was manually acquired and labelled.

Results

Data acquisition and preprocessing. For the image acquisition it was found that it is very important to keep the objects of interest in focus, such that their image exhibits sharp edges, which helps for the object classification.

Electrical appliance detection in orthophotos

The results of the ElecDetect module on the acquired datasets are shown in Table 9. It can be seen that the detection rate for Nygade1001 is quite good if the class labels are not taken into account, due to the similarity of switches and sockets in this room. Furthermore, there is a number of false negatives due to the fact that there was a large

Electrical wiring hypothesis Different approaches in the grammar have been evaluated against the ground truth, see Figure 67. The ground truth (row 1) on top shows that the power lines are always connected to the ceiling and running to the central power distribution from there.

Name	Sockets			Switches			Combined		
	M	FP	FN	M	FP	FN	M	FP	FN
Nygade1001	2	1	26	3	17	5	19	4	17

Table 9: Electrical appliance recognition, in terms of matches (M), false positives (FP) and false negatives (FN) per trained class. A pair of detection and groundtruth symbols is matched if they have 90% overlapping area. Combined means that the evaluation does only count correct matched positions, and not class labels.

Name	OrthoGen [s]	ElecDetect [s]	WireGen [s]
Nygade1001	2.4	575.2	0.44
Nygade1002	3.0	952.6	0.36
Nygade1005-1006	8.7	2157.1	0.58
LysLab03	4.6	1309.7	0.48
Nygade037	3.1	899.5	0.37
Bygade72-2nd	59.1	13044.6	10.3

Table 10: Wall clock timings for the RISE modules.

Using a context free grammar (row 2 and 3) shows a hypothesis that connects all endpoints, but as the grammar is context free, the elements that lie outside of an installation zone suggested by the grammar (row 2) are connected in an unintuitive fashion.

Using a context sensitive grammar (row 4 and 5) yields installation zones when the data (columns of sockets/switches) also suggests an installation zone, which yields a more natural configuration.

The bottom rows (6 and 7) use an updated grammar that does not exhibit the installation zone on bottom, which yields a result that is very similar to the ground truth.

Processing Time This table shows the measured execution time for each component. Timings were taken on a Core i7-4930K with 3.40GHz and 6 cores (see also Table 10). It can be seen that OrthoGen and WireGen modules are in the range of seconds for the smaller datasets, while ElecDetect consumes considerably more time (about 15 to 30min for the smaller datasets and 3.6 hrs for the larger dataset). The reason is that ElecDetect performs an exhaustive search over the whole orthophoto range, a sliding window that tests all pixel positions for the appearance of a trained class.



Figure 67: Comparing several grammar approaches in the Nygade1001 dataset using ground truth detections (row 1): context free (row 2-3), context sensitive (row 4-5) and context sensitive without installation zones on bottom (row 6-7).

Conclusion

The RISE tool shows very good results upon datasets, when high resolution and high quality images and well reconstructed 3d geometry exists. With this prerequisites the detection rate of sockets and switches is good and the wiregen component generates in its last version good assumptions on the routes, that cables are laid in the wall. the goal of this DURARK task to detect (almost) invisible elements in 3d scans was achieved. There is still room for improvement on a technical level (e.g. improve the detections by using a more recent computer vision algorithm), but the overall approach is able to scale in the foreseeable future.

Stakeholders, to whom the results were presented, commented, that the tool is sufficiently accurate for the use in a stakeholder environment in order to estimate for instance the length of cables, but especially to detect very small features in large spaces. They suggested to look for further elements, as for instance the heads of sprinkler systems, which can currently only be checked manually or for way signs or posters - elements, which reveal their meaning only through the symbolic meaning, but not through the shape. The herein developed approach presents an alternative to current technologies for the search of elements in Point Clouds. A prerequisite for a success in a commercial setting is a better quality of the retrieved images. Mobile scanning platforms, which integrate image and LIDAR scan, such as the Pegasus:Backpack by LEICA⁶¹ Leica, and better HDRI image acquisition devices, such as the iSTAR by NCtech⁶² are possible future platforms for the further development of the approach.

⁶¹<http://leica-geosystems.com/products/mobile-sensor-platforms/capture-platforms/leica-pegasus-backpack>

⁶²<https://www.nctechimaging.com/istar/>

8 Evaluation concerning Long Term Accessability

This section describes the evaluation activities concerning topics related to the Long Term Accessability of the DURAARK system and components developed for the preservation of three dimensional data. With the help of these evaluation activities the DURAARK project was able to verify, if the developed strategies are beneficent for the needs of the defined stakeholders and if 3D data could be successfully ingested into and retrieved from an existing digital preservation system with the help of the developed components.

The following topics have been evaluated:

- the long term archiving policies in Deliverable D6.3 presented through stakeholder interviews,
- the SIP Generator component and the ingest of a Rosetta compliant SIP into an existing digital preservation system,
- the schema compliance of the metadata extractors for the technical metadata ifcm and e57m.

The retrieval from the ingested data from the Rosetta system is not evaluated separately. The reason is that we show the successful ingestion and the generation of a valid, unique identifier for the respective data. Rosetta is providing an official web service API for the retrieval of the data based on this unique identifier. We consider this as default functionality supported by Rosetta and do not evaluate the functionality explicitly here. The Rosetta Connector component implements the functionality to retrieve data from Rosetta based on the unique identifier (SIP ID, see Section [8.2](#)).

8.1 Long term archiving policies

As detailed described in Section 5 of Deliverable D6.3 long term archiving policies are required input factors for a sustainable and successful preservation planning process within an OAIS archive. The goal of this section is on show how the presented policies are relevant for possible stakeholders.

Aspects evaluated

The preservation policies developed in DURAARK are based upon the SCAPE Policy Framework⁶³, which was a major outcome of the EU funded project SCAPE. Event though the policies are strongly connected to the work of the SCAPE project, they consider the results of the investigations which have been conducted for the work in Deliverable D7.3. Hence, the DURAARK project has developed a Guidance Policy "DURAARK Preservation" with its consisting Preservation Procedure Policies:

- Authenticity
- Metadata
- Preservation Planning
- Processes and Provenance
- Standards
- Access
- Rights

Central questions and identified risks of these policies have been formulated as a base for the evaluation. The stakeholder questionnaire reuses the questions which are intended to foster discussion about particular Preservation Procedure Policies.

Method and Evaluation activities

The DURAARK preservation policies was evaluated with the help of stakeholder surveys. Within the policies the needs and possible risks of specific tasks have been identified from an established Long Term Archiving perspective. The DURAARK policies were the base of the stakeholder questionnaire. The answers provide an indication, whether the needs of the stakeholders are covered in the DURAARK policies.

In total, ten questionnaires were filled out and evaluated by DURAARK. Since not every questionnaire was filled out completely some answers only feature a subset of the questioned stakeholders.

⁶³*<http://wiki.opf-labs.org/display/SP/SCAPE+Policy+Framework>

The questionnaire consists of seven parts: one general part and parts for each developed Preservation Procedure Policy as mentioned above. However, the policies "Preservation Planning" and "Processes and Provenance" were handled as "Long term preservation" within the questionnaire. The part "Rights" of the questionnaire is additional and not covered by the DURAARK preservation policies.

Results

In the following paragraph, an overview of the results are presented, focusing on the question, if the developed DURAARK policies fit the needs of the stakeholders. Hence, not every question will be presented.

General The general question which areas are of importance inside a specific organization has shown that it is really heterogeneous and a general answer is not easy at all.

<i>Authenticity</i>	2,8
<i>Rights</i>	3,2
<i>Metadata</i>	2,9
<i>Long Term Preservation</i>	2,7
<i>Access</i>	3,2
<i>Standards</i>	3,2

Table 11: Spreading of importance

Table 11 shows an overview about the most important fields within the organizations. This is directly based upon the answers, where the stakeholders were asked to say what is important (=1) and what is not important (=6). The average was calculated and is presented in the table. Even though 'Long Term Preservation', 'Metadata' and 'Authenticity' were the most important tasks for the institutions, the other fields are important likewise with a gap of 0,3 to 'Metadata'.

Hence, the developed DURAARK policies seem to fit the general needs of the stakeholders.

Additionally it was asked if there are already policies in place.

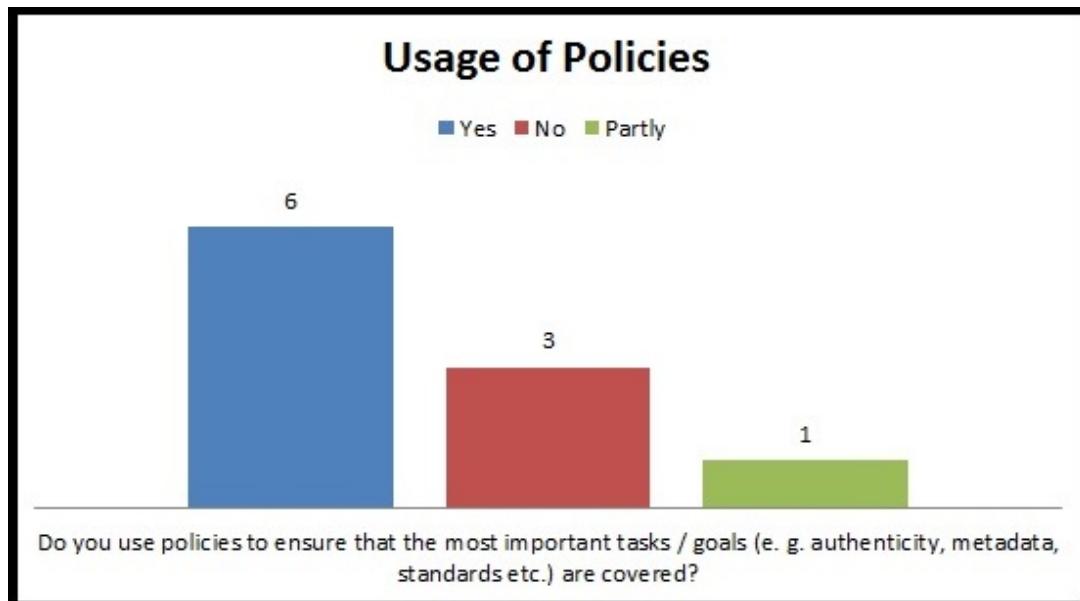


Figure 68: Policies within the institutions

Figure 68 shows that a majority of the asked stakeholders have policies in place and are probably open for the DURAARK preservation policies as well to reach fulfill their goals and tasks.

Authenticity / Integrity

As described above, the authenticity and integrity was in average almost the most important issue for the stakeholders. They were asked, how trustworthiness within the organization and for the user is established with regards on the authenticity and integrity on the preserved objects:

- *8 of 10 stakeholders* answered to establish transparent workflows rights and metadata to establish trustworthiness and to increase the reliability
- *7 of 10 stakeholders* answered that standards are used for reaching this goals
- *3 of 10 stakeholders* track every manipulation, edition, deletion of the data / metadata and make them visible to the user
- *2 of 10 stakeholders* are striving for certification to establish trustworthiness

- *1 of 10 stakeholders* uses policies for this goal

It has become very clear, that the stakeholders already try to establish trustworthiness and increase the reliability on the user's side. 80 respectively 70 percent are using recommended approaches like transparent workflows and standards for data and metadata. With 30 percent stakeholders tracking every change which has been conducted on an object a central requirement of sustainable and authentic preservation is also in place. However, only one stakeholder uses policies for regulating the authenticity of the preserved materials.

According to this results and the fact that authenticity of the preserved material is of high importance for the stakeholders, the DURAARK policy on "Authenticity" seems to fit the needs of the stakeholders.

Rights

Right on digital material are often highly specific, as they are dependent of the goals of an institution, the digital data itself and the heterogeneous kind of rights.

<i>Copyrights</i>	2,25
<i>Access rights</i>	1,4
<i>Licences</i>	1,6

Table 12: Importance of kind of rights

The stakeholders were asked, which kind of rights are the most important (1=important, 2=less important, 3=unimportant), respectively, which have to be considered the most. Table 12 shows a clear tendency as the access rights seem to be the most important type of rights on the stakeholder's side. Hence, it is not surprising that licences are also of high relevance, as they are connected with the access rights. The DURAARK policy on "Rights" focuses strongly on the access rights of architectural 3D data and fits the needs of the stakeholders, too.

Metadata

Metadata were considered above as one of the most important areas which has to be covered within the stakeholder's organizations. The stakeholders were asked, which kind

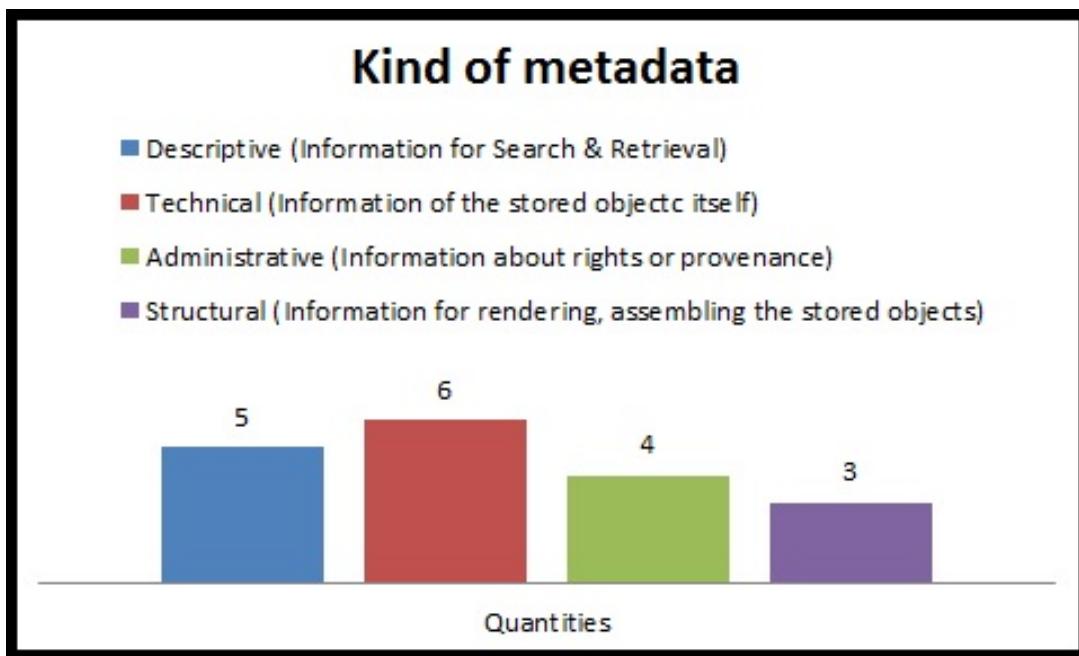


Figure 69: Used kinds of metadata

of metadata is captured within their organization. Even though not every stakeholder answered the question, it becomes clear, that technical and descriptive metadata are captured within the organizations as shown in Figure 69. For digital preservation every kind of metadata is necessary and required, but technical and descriptive information is of high significance. Without descriptive and technical information a sustainable long term preservation is not feasible.

The stakeholders were additionally asked, if at least a minimal set of metadata is required and if persistent identifiers are in place for search and retrieval:

- *6 of 10 stakeholders* require a minimal set of metadata
- *2 of 10 stakeholders* have persistent identifiers in place

The investigation has shown, that metadata is certainly of importance within the organization. But for instance, digital preservation is only hardly feasible if a persistent identifier is missing, as it ensures the retrieval over the long term if systems and infrastructure is altered. Since the stakeholders seem to be really familiar with metadata, the DURAARK policy on "Metadata" is surely a useful approach for metadata adjustments.

Long Term Preservation

In the beginning of the questionnaire, long term preservation was considered as the most important task within the stakeholder's institutions. The questionnaire tried to identify, if established and recommended functions are in place to ensure the availability of the data over the long term:

- *1 of 10 stakeholders* has services establish to monitor the change of technology with regards to access and renderability of the stored data
- *2 of 10 stakeholders* checks if the visual appearance and the content of a file looks the same if the file is open with another (=higher or newer) version of the program.

Given the fact that long term preservation was considered as most important, these quantities are quiet surprising. It can be assumed, that long term preservation is indeed of high relevance inside the organizations but implementations are not yet in place. However, the questions about long term preservation have also shown, that the stakeholders are familiar with several services, like redundant storage, backup strategies and system monitoring to enable at least a secure and sustainable bit preservation of the preserved material.

The DURAARK approach can help to establish long term preservation functionalities but first a deeper understanding for digital preservation probably have to be established.

Access

Access and reuse of preserved material is for the most organizations a core goal. Hence, it was not rated with highest importance above but of average importance indeed. Some questions were conducted, how the organizations offer their material for the users:

- *6 of 10 stakeholders* offer the data via web interface
- *5 of 10 stakeholders* make the data available via a client side software
- *3 of 10 stakeholders* provide the data as a direct download
- *1 of 10 stakeholders* uses a database for access

Some of the stakeholders offer different approaches for making their data available. However, it can be stated that nearly every stakeholder makes the data available in some way or other. With regards to independent understanding of the data, the 5 of 10 stakeholders offer additional descriptive or structural information.

Again, the DURAARK policy seems to meet the requirements of the stakeholders. With

the help of this particular policy, the process of access can be deepen and arranged in a more formalized way. Especially with regards to representation information or the transformation of access copies.

Standards / Certification

For some questioned stakeholders standards and certificates were the most important goals which have to be covered within their organization. The DURAARK project was interested, which kind of Standards are of high importance:

- *7 of 10 stakeholders* consider file format standards as important
- *5 of 10 stakeholders* consider metadata standards as important
- *4 of 10 stakeholders* consider reference models as important
- *3 of 10 stakeholders* consider storage standards as important
- *1 of 10 stakeholders* consider other standards as important
- *1 of 10 stakeholders* doesn't consider standards as important

Hence, the file format standardization plays a great role within the stakeholder's organization. The same applies for metadata at least for the half of the questioned stakeholders. And also reference models seem to play a role in these organizations. Additionally, the stakeholders were asked, if certification is also of significance in their institution but only one stakeholder affirmed.

Even though standards are accepted within the questioned organizations, the DURAARK policies are at least helpful to review the own standarization approaches.

Conclusion

The survey has clearly shown, that even within ten questioned stakeholders a broad variety of needs, concerns and requirements exist. The DURAARK Policies seem to fit the needs of the stakeholders. In most cases they can help stakeholders to formalize activities, which are already in place, or even to implement new activities, like those regarding long term preservation.

8.2 SIP Generator / Connection to Rosetta Ingest

The TIB uses Rosetta by Ex Libris as digital preservation system for their digital material. The DURAARK approach concentrates on the delivery of 3D data from the WorkbenchUI into an existing system. Hence, it was necessary to create a Rosetta-compliant SIP structure and deliver it with the official Rosetta web services.

The goal of this section is to give a comprehensible overview of the developed SIP Generator, its supported input structure and scalability and the delivery functionality to ingest the material into Rosetta.

Aspects evaluated

Three different aspects have been evaluated:

- **SIP Generator:** The generation of a Rosetta-compliant SIP structure with one or multiple representations and embedded technical and descriptive metadata as SourceMD.
- **Rosetta Connector:** The transfer of the generated Rosetta SIP into an existing Rosetta DPS.
- **Data ingestion:** Verification if the material is correctly ingested with its multiple representations and SourceMD

Method

As described in Section 2 of the Deliverable D6.2 an AIP can consist of more than just one representation. For the purpose of evaluation, heterogeneous Rosetta SIPs have been packaged, with having just one master representation and also master and derivative representations. Along with this data, descriptive metadata - buildm - and technical metadata - ifcm for IFC master files, e57m for E57 master files - is provided in a separate folder. It will be checked, if the SIP Generator transform this input structure into a Rosetta-compliant one.

In a next step the Rosetta Connector is uploading the SIP to the Rosetta DSP and triggers its ingestion via Rosetta's Deposit web service ⁶⁴.

After the ingestion process is finished by Rosetta we check if all representations have been delivered correctly and if all provided metadata are successfully displayed as SourceMD.

⁶⁴<http://rosetta.develop.lza.tib.eu/dpsws/deposit/DepositWebServices?wsdl>

Datasets and Evaluation Activities

During the conception and development of the SIP Generator component it was clear that the SIP generator should be able to process different input structures. Input structures can consist of one or multiple master file(s), where each master file can have a derivative file. A derivative file can be generated by geometric enrichment components in the WorkbenchUI and is using the respective master file as input (e.g. the IFC reconstruction component generates an IFC file as derivative of the respective E57 point cloud file). Hence, it is possible to ingest just one master file - no matter if it is an IFC or an E57 file -, a master file and a corresponding derivative copy or several master files with or without corresponding derivative copies. Along with the files, descriptive metadata as buildm and technical metadata as ifcm or e57m is processed. The SIP generator scales furthermore regarding the file size of the ingested files (see example file list below).

For evaluation the WorkbenchUI was used to produced four different SIP packages:

- **single_e57** consisting of one E57 file as master representation its corresponding technical and descriptive metadata (*size: 270.415.422 Bytes*)
- **single_ifc** consisting of one IFC file as master representation, its corresponding technical and descriptive metadata (*size: 17.588 Bytes*)
- **one_e57_one_derivative** consisting of one E57 file as master representation, its corresponding technical and descriptive metadata and one ifc file as derivative representation (*size: 538.460.332 Bytes*)
- **two_master_one_derivative** consisting two different intellectual entities: one ifc file as master representation, its corresponding technical and descriptive metadata; one E57 file as master representation, its corresponding technical and descriptive metadata and one ifc file as derivative representation (*size: 281.178.282 Bytes*)

As shown, for instance, in Figure 70, there is one E57 file as a master and one IFC file as a derivative representation present before processing through the SIP generator. Each of those in an appropriate folder. The metadata is located in the folder "sourcemd", to be embedded as SourceMD in the Rosetta-compliant METS file.

```

<---ie_id1
    +---derivative_copy
        |   Nygade_Scan1005-1006_RECONSTRUCTED.ifc
    |
    +---master
        |   Nygade_Scan1005-1006.e57
    |
    \---sourcemd
        buildm.xml
        e57m.xml

```

Figure 70: One master file, one derivative copy and their corresponding metadata files

Results

All in 8.2 presented datasets⁶⁵ were processed by the SIP Generator. The input folder structure was transformed into a Rosetta-compliant one. As presented in Figure 71, the folders with the representations ("‘derivative copy’" and "‘master") have been embedded into the new created folder "‘streams’" which itself is part of the also newly created folder "‘content’". The folder "‘sourceMD’" with the metadata has been removed. The added METS file "‘ie1.xml’" was generated according to the Rosetta profile⁶⁶ and the original metadata was embedded into this file as "‘sourceMD’" part. Additionally, the "‘dc.xml’" was added for presentation and indexing in the Rosetta System.

```

<---ie_id1
    dc.xml
    |
    \---content
        ie1.xml
        |
        \---streams
            +---derivative_copy
                Nygade_Scan1005-1006_RECONSTRUCTED.ifc
            |
            \---master
                Nygade_Scan1005-1006.e57

```

Figure 71: Rosetta-compliant SIP structure with corresponding METS file

Ex Libris offers web services⁶⁷ to deliver a package into an existing Rosetta DPS without direct access to the system.

In Figure 72 the parameters for a successful deposit into the system are shown, for instance the ID of the used material flow, the name of the source directory for collecting the data

⁶⁵Due to lack of space only one folder will be presented in pictures in the following.

⁶⁶<http://www.loc.gov/standards/mets/profiles/00000042.xml>

⁶⁷<https://developers.exlibrisgroup.com/rosetta/apis>

or the ID of the specific producer. The service also indicates if the deposit was successful and the SIP ID of the particular SIP package. This SIP ID will be permanently stored along the submitted package and is therefore a unique identifier inside the system (also, we use the SIP ID as identifier for the archived data within the DURAARK system, e.g., it is stored in the respective buildm entry within the SDAS for later retrieval of the data).

```
<ser:deposit_result xmlns:ser="http://www.exlibrisgroup.com/xsd/dps/deposit/service">
  <ser:is_error>false</ser:is_error>
  <ser:sip_id>34221</ser:sip_id>
  <ser:deposit_activity_id>40576</ser:deposit_activity_id>
  <ser:user_params>
    pdsHandle=1211201515532515142421882958866801,
    materialFlowId=36411632,
    subDirectoryName=tmp/aa96eb64-8e73-4eaa-bb9a-81e362b34f2a/ie_id1,
    producerId=36384081,
    depositSetId=1
  </ser:user_params>
  <ser:creation_date>Thu Nov 12 15:53:25 CET 2015</ser:creation_date>
</ser:deposit_result>
```

Figure 72: web service response

Inside the Rosetta system the submitted content can be accessed by searching for the SIP ID. Figure 73 shows that the specific SIP was submitted successfully into Rosetta, assigned to its specific workflow and producer ("TIB DURAARK Automatisch" and "TIB DURAARK"). Also the file name ("Nygade_Scan1005-1006.e57") is listed as title of the specific package.

SIP ID	34221	Producer	TIB DURAARK	
Deposit Activity ID	40576	Material Flow Name	TIB DURAARK Automatisch	
Title	Nygade_Scan1005-1006.e57	Selected Access Rights	Keine Beschränkung	
Issue #	-	Assign To	panitzm	
Content List	Metadata	Notes	History	
filter [All]			1 - 1 of 1 Re	
Title	PID	Status	CMS ID	Actions
Nygade_Scan1005-1006.e57	IE672611			View Edit IE

Figure 73: Overview over the SIP inside Rosetta

The described package consists of a master and a derivative representation. The shown representations were successfully delivered and presented inside the DURAARK DPS as shown in Figure 74. Every representation and file - as a representation can consist of several files as well - is added a unique ID, like "FL672613" or "REP672614".

As described above, the original metadata files buildm and e57m respectively ifcm were removed and their content was copied into the METS file "ie1.xml". Within Rosetta this

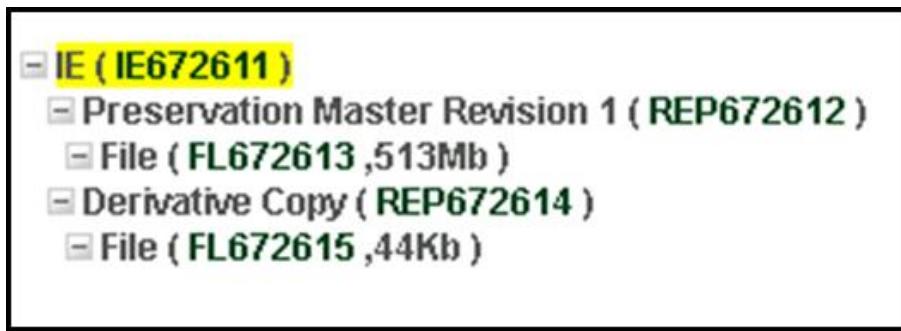


Figure 74: Different representations of the SIP

metadata can be accessed in the metadata section of the corresponding package. Figure 75 shows that the metadata - in this case e57m - is part of the package. It can be accessed or edited, if necessary.

MID	57923	Metadata Type	source:other
Created by	TIB-DURAARK-Auto	Creation Date	12.11.15
Updated by		Update Date	12.11.15
<pre><e57m xmlns:mets="http://www.loc.gov/METS/"> <E57root> <guid>(9CAD8612-B6CF-4111-A71-C90E21F2E22D)</guid> <version>1.0</version> <creation_datetime>2015-02-19T08:54:56.896Z</creation_datetime> <coordinate_metadata>Not in use.</coordinate_metadata> <scan_count>2</scan_count> <image_count>0</image_count> <scan_size>2</scan_size> <image_size>0</image_size> </E57root> <e57scan></pre>			

Figure 75: SourceMD of a particular SIP package

8.3 Metadata extractor (IFC, E57)

Aspects evaluated

This evaluation shows that the technical metadata (XML format) produced by the metadata extractor components for IFC and E57 file is valid with respect to the ifcm and e57m schema developed and expected by the Rosetta system.

Method

Selected datasets will be ingested into the Rosetta system, where the technical metadata of the IFC and E57 files was extracted as XML files by the developed metadata components. The success criteria is the successful ingestion into the Rosetta system.

Datasets and Evaluation Activities

For the evaluation the same datasets as in the previous Section 8.2 have been used. All data have been processed within the WorbenchUI. The technical metadata was implemented as SourceMD in the METS file during the SIP generation process.

Results

The ingestion of the technical metadata into the Rosetta system was, as also described in Section 8.2 respectively shown in Figure 75, successful.

9 Conclusion

The evaluation activities of the DURAARK consortium succeeded in covering all research output from all Workpackages. The project succeeded furthermore in the efforts to establish a common evaluation framework, that could be generally applied over the output of all involved disciplines and provide a common measure for the projects outcomes and the relation to the identified stakeholder groups.

The consortium agreed on five scientific criteria for the evaluation:

- Relevance
- Utility - Integratability
- Precision
- Usability
- Quality

The degree to which each of these could be applied in the evaluation of the research output was naturally strongly depended on the state and nature of the research output. Some of the evaluated components, such as the Focused Crawler (Section 6.1), are so far ahead of the current stakeholder practice, that they cannot be evaluated in their environment, while others, such as the Geometric Enrichment tools, can directly work on datasets and in environments of stakeholder. This doesn't mean, that the research is not relevant for the stakeholders, but solely, that the evaluation methods and techniques have to be adapted. Stakeholders confirmed the relevance for all components, no matter how far into the future they are pointing.

A common challenge was the interaction with stakeholders for the evaluation. The number of available experts in the highly specialised and foreword looking areas, that DURAARK touches upon, is limited and the existing expert under high demand. Practitioners need furthermore an incentive to partake in evaluation activities and invest the time to contextualise the necessarily abstract research presented to them in the evaluation session. This incentive is stronger, when the research result is tangible. This was very well visible in the evaluation of the Geometric Enrichment components, were tools were finally available and processes developed, which looked promising to stakeholders. This gave willingly their models - a task which was enormously difficult in the years before. The efforts to

gain this tangibility are however quite high - especially, when the research project is of a strategic nature, as in a STREP project. The efforts to test i.e. the integratability of the components were high. However these activities in itself led to new insights, software libraries, such as the newly written e57 .Net libraries, and impulses, that contribute to the application of the research results in further collaborations with stakeholder, as reported on in D8.7.

The evaluation activities contributed to research practice, as in some cases new evaluation approaches and metrics had to be developed. The description of the deviation between different three-dimensional representation of a building, where a "ground truth" is hard to define is such a case. The DURAARK consortium developed here metrics to compare the results across many tested datasets in automated and semi-automated ways and provide means to draw the right conclusions for further developments.

The conclusions taken from the evaluation activities on the system and component level are described in the related sections in the document. They have however in common that they could feed into the development process of each component and the relation to the overall. The evaluation activities can hence not solely be understood as a measure, whether the aims of DURAARK were reached - which could actually be confirmed, but as a mean to generate criteria to decide, impulses and inspiration within the research phase and beyond.

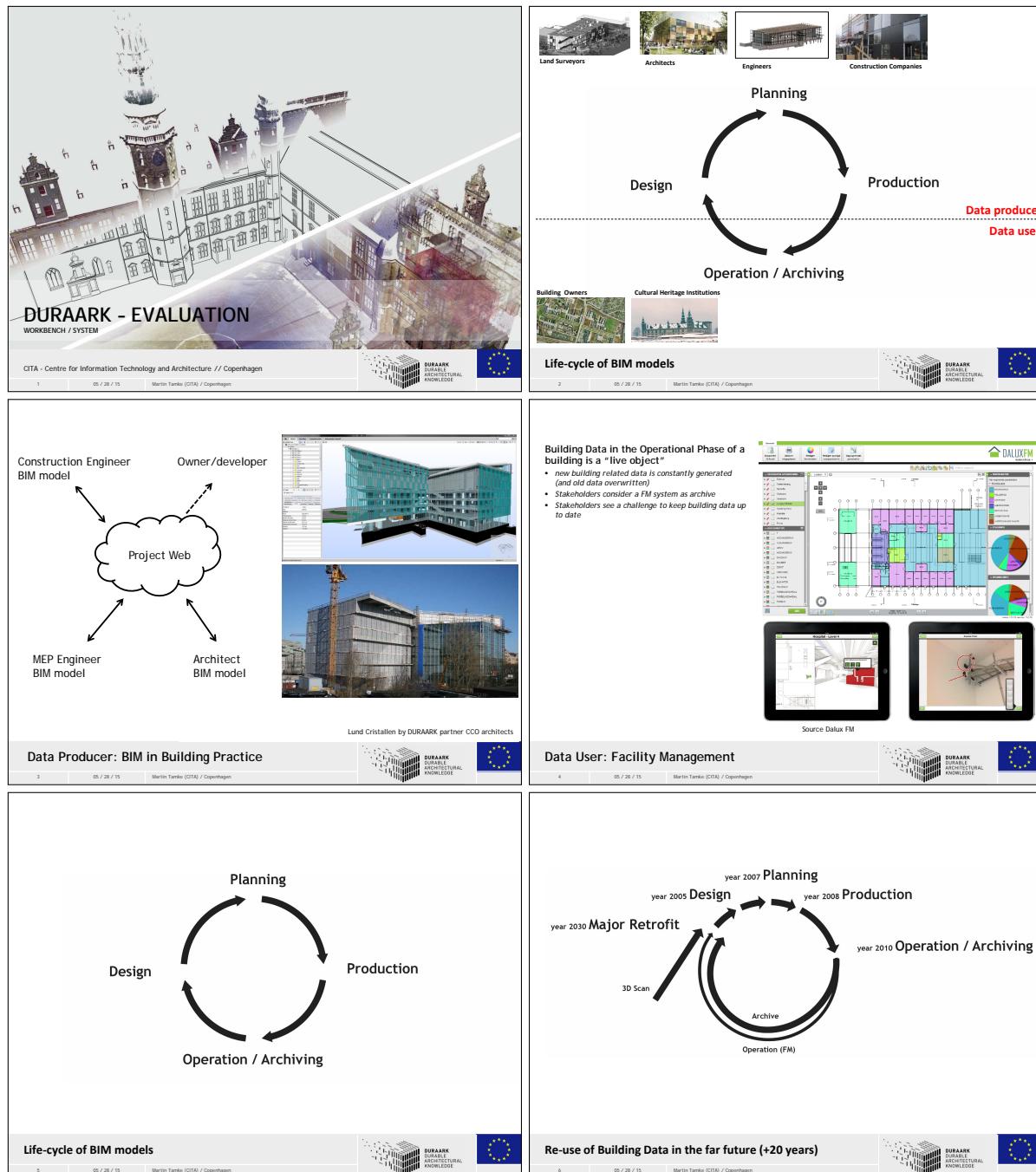
References

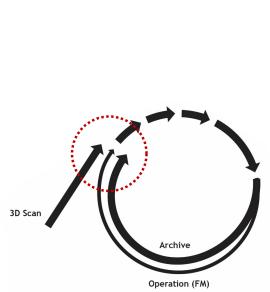
- [1] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1):107–117, 1998.
- [2] Soumen Chakrabarti, Kunal Punera, and Mallela Subramanyam. Accelerated focused crawling through online relevance feedback. In *Proceedings of the 11th International Conference on World Wide Web, WWW '02*, pages 148–159, New York, NY, USA, 2002. ACM.
- [3] Randy Deutsch. *Data-driven design and construction: 25 strategies for capturing, analyzing and applying building data*. Wiley, 1 edition.
- [4] Besnik Fetahu, Ujwal Gadiraju, and Stefan Dietze. Improving entity retrieval on structured data. In *The Semantic Web - ISWC 2015 - 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part I*, pages 474–491, 2015.
- [5] Ujwal Gadiraju, Ricardo Kawase, and Stefan Dietze. A taxonomy of microtasks on the web. In *Proceedings of the 25th ACM conference on Hypertext and social media*, pages 218–223. ACM, 2014.
- [6] Ujwal Gadiraju, Ricardo Kawase, Stefan Dietze, and Gianluca Demartini. Understanding malicious behaviour in crowdsourcing platforms: The case of online surveys. In *Proceedings of CHI'15*, 2015.
- [7] N. Hichri, C. Stefani, L. De Luca, P. Veron, and G. Hamon. From point cloud to bim: A survey of existing approaches. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5/W2:343–348, 2013.
- [8] J.-F. Hullo, G. Thibault, and C. Boucheny. Advances in Multi-Sensor Scanning and Visualization of Complex Plants: the Utmost Case of a Reactor Building. *ISPRS -*

International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, pages 163–169, February 2015.

1 Appendix

1.1 DURAARK WorkbenchUI Test - Contextualisation





- How to maintain access to information?
- How to trust the information?
- How to use the information?

Building and Building Data Lifecycle

7 05 / 28 / 15 Martin Tamke (CITA) / Copenhagen DURAARK DURABLE ARCHITECTURAL KNOWLEDGE EU



Point Cloud Data BIM 2D Data Geo Data WWW (Linked Data Cloud)

PLAN 3D Dataset provided by Plan3D / Berlin

Motivation: Enrich Architectural data, preserve meaning

8 05 / 28 / 15 Martin Tamke (CITA) / Copenhagen DURAARK DURABLE ARCHITECTURAL KNOWLEDGE EU

How to maintain access to information?

semantic preservation	conceptual object	authenticity, interpretability "How to understand/ interpret the data?"
logical preservation	logical object	logical preservation "How to open/render the file?"
bit preservation	physical object	bit preservation "How to keep the 1s and 0s?"

Long Term Archiving- 3 Layers of preservation

9 05 / 28 / 15 Martin Tamke (CITA) / Copenhagen DURAARK DURABLE ARCHITECTURAL KNOWLEDGE EU



DURAARK
DURABLE
ARCHITECTURAL
KNOWLEDGE

www.DURAARK.eu

10 05 / 28 / 15 Martin Tamke (CITA) / Copenhagen DURAARK DURABLE ARCHITECTURAL KNOWLEDGE EU

DURAARK Consortium

11 05 / 28 / 15 Martin Tamke (CITA) / Copenhagen DURAARK DURABLE ARCHITECTURAL KNOWLEDGE EU

How to maintain access to information?

bit preservation	physical object	bit preservation "How to keep the 1s and 0s?"
------------------	-----------------	--

- Existing Long term Archiving systems (Rosetta)

Long Term Archiving- 3 Layers of preservation

12 05 / 28 / 15 Martin Tamke (CITA) / Copenhagen DURAARK DURABLE ARCHITECTURAL KNOWLEDGE EU

How to maintain access to information?		How to maintain access to information?	
	<p>logical preservation "How to open/render the file?"</p> <ul style="list-style-type: none">• Use well documented open standards• Validate the files whether they follow standards		<p>authenticity, interpretability "How to understand/interpret the data?"</p> <p>The hard question:</p> <ul style="list-style-type: none">• Keep data in use - Create added value through multiple use of data for many• Be able to find the data - Link data to other data• Create linkage (semantic information) to maintain understanding• Keep data up to date:<ul style="list-style-type: none">• to new file standards• To changes in buildings
Long Term Archiving- 3 Layers of preservation		Long Term Archiving- 3 Layers of preservation	
<ul style="list-style-type: none">• Longterm Archiving system• Archiving Policies• Generate Semantic Information on Textual and geometric level• Search & Retrieval			
<ul style="list-style-type: none">• Tools to Evaluate information			
<ul style="list-style-type: none">• Integration of System and archived Information into existing workflows			
DURAARK Approach		DURAARK workbench	
<ul style="list-style-type: none">• Webbased• Modular• Adaptable <p>Integratable into existing workflows and systems</p>			
DURAARK Longterm Archiving Approach		DURAARK Ingest	

<p>Metadata IFC buildM schema</p> <p>Machinereadability Semantic Enrichment - Geometry</p> <p>Linkage Semantic Enrichment - WWW</p> <p>DURAARK Ingest</p> <p>19 05 / 28 / 15 Martin Tamke (CITA) / Copenhagen</p>	<p>RDF Subject → RDF Predicate → RDF Object</p> <p>New... that should clear up a few things around here © Horrocks, Oxford University</p> <p>Search: Semantic Web technology</p> <p>20 05 / 28 / 15 Martin Tamke (CITA) / Copenhagen</p>
<p>Metadata: Search for Geolocation</p> <p>Automated Reconstruction of BIM model from Point Cloud: Search for elements</p> <p>Point Cloud and BIM model: Detection Differences</p> <p>Scan and Image: Find almost invisible thing - detection of electrical appliances in walls</p> <p>DURAARK Search</p> <p>21 05 / 28 / 15 Martin Tamke (CITA) / Copenhagen</p>	<p>Geo Data</p> <p>Linked Data Cloud</p> <p>Explore Context Information</p> <p>DURAARK Semantic Digital Archive: Find related relevant data</p> <p>22 05 / 28 / 15 Martin Tamke (CITA) / Copenhagen</p>
<p>Semantic Digital Archive (SDA) consists of three components:</p> <ol style="list-style-type: none"> 1. Meta Data Extraction and Semantic Enrichment <ul style="list-style-type: none"> Industry Foundation Classes (IFC), STEP Physical File Format (SPFF) created with native BIM Software (e.g. ArchiCAD, Nemetschek, Revit) <p>Meta Data Extraction</p> <p>23 05 / 28 / 15 Martin Tamke (CITA) / Copenhagen</p>	<p>1. Semantic Digital Observatory (SDO) crawling</p> <p>SDO: Crawling relevant data sets for enrichment</p> <p>24 05 / 28 / 15 Martin Tamke (CITA) / Copenhagen</p>

Connectedness

Specific to the Built Environment

Linkage to Existing models and vocabularies

Semantic Digital Archive (SDA) consists of three components:

1. Meta Data Extraction and Semantic Enrichment
 - Extract meta data from E57 and IFC files submitted to the archive.
 - Enrich by curator and automated methods
2. Semantic Digital Observatory (SDO)
 - Discover relevant data sets for semantic enrichment
 - Cluster similar data based on initial seed list

SDO: Clustering datasets

Semantic Digital Archive (SDA) consists of three components:

1. Meta Data Extraction and Semantic Enrichment
2. Semantic Digital Observatory (SDO)
3. Semantic Digital Archive Storage (SDAS)
 - Store meta data of archived content and expose as Linked Data

http://mimas.cgi.tugraz.at/search

Search in Building Data - Exposed archived meta data

DURAARK: Integration into existing Workflows

DURAARK Workbench: Integration

DURAARK
DURABLE
ARCHITECTURAL
KNOWLEDGE

www.duraark.eu
<https://github.com/DURAARK>