D6.3 Report on sample Preservation Planning for 3D objects

DURAARK

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

Date: 2016-01-31
Version 1.0
Document id. : duraark/2016/D.6.3/v1.0

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 – Kunstakademiets Arkitektskole [DK] LTU – Lulea 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/2016/D.6.3
Title of document	:	D6.3 Report on sample Preservation Planning for 3D objects
Deliverable type	:	Report
Contractual date of delivery	:	2016-01-31
Actual date of delivery	:	2016-01-31
Lead beneficiary	:	LUH
Author(s)	:	Michael Panitz <michael.panitz@tib.eu> (LUH)
	:	Ujwal Gadiraju <gadiraju@l3s.de> (LUH)
Responsible editor(s)	:	Michael Panitz <Michael.Panitz@tib.eu> (LUH) .
Quality assessor(s)	:	Jakob Beetz <J.Beetz@tue.nl> (TUE), Martin Tamke <Martin.Tamke@kadk.dk> (CITA).
Approval of this deliverable	:	Stefan Dietze <dietze@L3S.de> (LUH) – Project Coordinator
Distribution	:	Public
Keywords list	:	OAIS integration, digital preservation, preservation planning, semantic preservation, policies

Executive Summary

This deliverable describes a sample preservation planning of 3D objects through the example of migration of IFC files. It provides a workflow for preservation planning and features an in-depth discussion on requirements and constraints for preservation planning, such as significant properties or preservation policies. A sample preservation plan was conducted based on the defined workflow and requirements. The in Deliverable D6.6.1 identified best practices solutions are also taken into account within this deliverable. Additionally, policies are described which formalize the interaction of DURAARK's SDA with a OAIS-compliant archive. A exemplary ingestion of RDF snapshots of the SDA is provided at the end of this deliverable.

Table of Contents

1 Introduction . . . 5
1.1 Implementation level . . . 6
1.2 Methodology . . . 7
1.3 Risks . . . 8
1.4 Technical Decisions . . . 11
2 Use cases . . . 12
2.1 UC1: Deposit 3D architectural objects . . . 12
2.2 UC3: Maintain Semantic Digital Archive . . . 13
3 Preservation Planning Process . . . 14
3.1 Prerequisites . . . 15
3.2 Plato Workflow . . . 16
3.3 DURAARK Preservation Workflow . . . 18
3.4 Conclusion . . . 23
4 Significant Properties . . . 24
4.1 Technical input factors . . . 24
4.2 Organizational input factors . . . 43
4.3 Stakeholder requirements . . . 45
5 Preservation Policies . . . 49
5.1 Policies as input factors . . . 49
5.2 SCAPE Preservation Policy Framework . . . 51
5.3 DURAARK Policies . . . 52
5.4 Institutional preservation policies . . . 78
6 DURAARK Sample Preservation Plan . . . 80
6.1 File Level - Migration to new IFC schema . . . 80
6.2 SDA Preservation . . . 90

7 Conclusion and DURAARK Objectives 97
7.1 Integration Scenarios 97
7.2 Impact and Outlook 98
7.3 Collaborations with other projects 98
References 101

A Appendix 102
A.1 SCAPE Guidance Policy Elements 102
A.1.1 'Authenticity' 102
A.1.2 'Bit Preservation' 103
A.1.3 'Functional Preservation' 103
A.1.4 'Digital Object' 103
A.1.5 'Metadata' 103
A.1.6 'Rights' 104
A.1.7 'Standards' 104
A.1.8 'Access' 104
A.1.9 'Organization' 104
A.1.10 'Audit and Certification' 105
A.2 Policy element template 106

1 Introduction

The Deliverable D6.2[1] has described in detail how to ingest and store architectural 3D data inside an existing digital preservation system (DPS). This was elaborated within the second prototype, as described in Deliverable D2.5[2], where the successful connection and deposit of architectural 3D data from the DURAARK WorkbenchUI into Rosetta was performed and described. Numerous gaps identified in the first WP6 Deliverable D6.6.1[3] were addressed were addressed in D6.2 including for instance the logical preservation or metadata for architectural 3D data. Identified issues regarding stakeholder requirements were addressed in D7.3[4]. In the initial Deliverable D6.6.1 gaps have also been identified with regards to organizational roles in digital preservation which have not been addressed yet:

- A common understanding of significant characteristics for architectural 3D data was missing due to the lack of technical metadata.
- No exemplary preservation plan for 3D architectural data was available.

Hence, this deliverable will focus on the Preservation Planning and its requirements for architectural 3D data. The OAIS reference model describes Preservation Planning as the

> *"functional entity which provides the services and functions for monitoring the environment of the OAIS and which provides recommendations and preservation plans to ensure that the information stored in the OAIS remains accessible to, and understandable by, and sufficiently usable by, the Designated Community over the Long Term, even if the original computing environment becomes obsolete."*[5]

This means that even if data is ingested successfully into an existing DPS, its accessibility, understandability and usability is not necessarily ensured in the long term. The developers and reviewers of the OAIS framework discovered the same difficulty as Preservation Planning was originally not a part of the first OAIS draft in May 1999. It was the 'Networked European Deposit Library' which required and suggested a specific entity inside the OAIS for ensuring the long term preservation, which was then entitled 'Preservation

[1] `http://duraark.eu/wp-content/uploads/2015/02/DURAARK_D6.2.pdf`
[2] `http://duraark.eu/wp-content/uploads/2015/08/DURAARK_D2_5_final.pdf`
[3] `http://duraark.eu/wp-content/uploads/2014/02/duraark_d6.6.1_final.pdf`
[4] `http://duraark.eu/wp-content/uploads/2015/02/DURAARK_D7.3.pdf`

Planning'.[4]
For ensuring long term preservation of architectural 3D data it is therefore of high importance for DURAARK to provide the preservation planning process for this specific kind of data.
DURAARK has developed a preservation planning process, described in Section 3, which enables the implementation of preservation planning and evaluation of preservation action.
Significant properties are often considered as the most important input factor when it comes to preservation planning. Section 4 features an in-depth discussion on the selection and types of significant properties, as they are not always connected to the digital object but can also stem from an organization.
Another often considered input factor are preservation policies which describe and formalize required processes but also constraints with regards to digital preservation. Section 5 discusses how policies do actually affect preservation planning and presents a set of developed DURAARK preservation policies.
In Section 6.1, the core of this deliverable, the sample preservation plan is introduced based on the previously defined requirements.

As also discussed in Deliverable D6.6.1, the Semantic Preservation of digital objects is necessary to ensure the independent understanding of the preserved content also over the long term. In the DURAARK project the Semantic Preservation is strongly connected to the semantic enrichment components SDA / SDO. Even though the SDA is referred to as 'Semantic Digital Archive' it offered no function for the long term preservation of its crawled and gathered semantic data so far. As required by the European Commission review board during the project review in March 2015, this deliverable describes developed policies for this purpose in Section 5.3.2 and presents in a first approach on how to preserve the SDA content also over the long term in Section 6.2.

1.1 Implementation level

As described in D2.5 and evaluated in D7.4 the ingestion of packages from the WorkbenchUI into Rosetta is already implemented and works fine. This includes the generation

of SIP packages according to the Rosetta METS profile[5] as well as the connection between WorkbenchUI and Rosetta with the help of the Rosetta Deposit Web Services[6].
As the evaluation in D7.4 has also shown, the metadata extractors for IFC and e57 files do not work in the Rosetta system due to operating system issues on TIB's side. Therefore the extraction of technical metadata takes place in the WorkbenchUI and its submission is handled as SourcMD inside the METS file.
The presented approaches for preservation planning and semantic preservation are conducted manually supported by the preservation planning functions that Rosetta provided. However, the preservation planning process follows an automatic assessment and has therefore not to be implemented.

1.2 Methodology

As described above, preservation planning and its required input factors as well as the policies and preservation for the SDA maintenance were still open gaps within the DURAARK project. The following methods have been used within this deliverable:

1. The developed DURAARK preservation planning process is based upon the work which has been conducted for the Preservation Planning Tool 'Plato'[7] within the PLANETS project[8] and the project 'SCAlable Preservation Environments' (SCAPE)[9]. In contrast to this approach, DURAARK has simplified the workflow and uses a manual assessment instead of using an automated tool to avoid additional technical infrastructure for Plato.

2. The selection of significant properties for IFC files is strongly based upon the recommendations and investigations which have been conducted in the project 'Investigating the Significant Properties of Electronic Content over Time' (InSPECT)[10] with regards to the object analysis.

[5] http://www.loc.gov/standards/mets/profiles/00000042.xml
[6] https://developers.exlibrisgroup.com/rosetta/apis/DepositWebServices
[7] http://www.ifs.tuwien.ac.at/dp/plato/intro/
[8] http://www.planets-project.eu/
[9] http://www.scape-project.eu/tools
[10] http://www.significantproperties.org.uk/

3. The presented policies have been developed upon the SCAPE Policy Framework[11] and its provided policy template. [12]
4. The sample preservation for IFC files uses the DURAARK preservation planning process as workflow developed in 1.) for the test run. The significant properties identified in 1.) and policies developed in 3.) function as input factors within this process and as measurable indicators for the success of the tested preservation action.

1.3 Risks

Risks associated with the addressed processes

#	Risk Description	Risk Assessment	Contingency Solution	Project Relevance
1	The identified significant properties do not apply user requirements.	**Impact**:Low **Probability**:High The within DURAARK conducted questionnaires have shown, how important the authenticity, integrity and access of architectural 3D data with regards to retrofitting or legal issues for the stakeholders is. Therefore the developed solutions are not only based on the DURAARK project partners but also on external input.	The project partners are in contact with possible stakeholders to verify their needs and to adjust this approach. Further test scenarios are also thinkable to broaden this specific selection approach.	Low (Observed)

[11] `http://wiki.opf-labs.org/display/SP/SCAPE+Policy+Framework`
[12] Both is presented in Annex A.1 and Annex A.2.

<table>
<tr><td colspan="5">Section 4 discusses furthermore, that significant properties are always dependent of the tasks and goals of an organization. Therefore a 100 percent objective approach is not feasible for the selection of significant properties. This particular part of this deliverable is rather a practical showcase than a generic approach for selection of significant properties.</td></tr>
<tr><td>2</td><td>The presented proof of concept for SDA snapshot preservation is not sufficient for sustainable long term preservation.</td><td>Impact:Middle
Probability:High
Within this deliverable only bit preservation for RDF snapshots of the SDA is provided. Even though the bit preservation is conducted within Rosetta, this might not be sufficient with regard to the long term availability and therefore to the semantic preservation of the architectural 3D data.</td><td>Further development is needed to ensure not only the bit preservation but also the logical preservation and the semantic preservation to ensure the the long term availability of the SDA.</td><td>High (Observed)</td></tr>
<tr><td colspan="5">Since the long term preservation of the SDA was not the major goal of the DURAARK project this risk was early considered and accepted. The offered solution has to be understand furthermore as a first approach on how to preserve the SDA and its content to enable for instance its re-population.</td></tr>
</table>

3	The used Software 'Autodesk Revit' is not an appropriate tool for IFC format migration.	**Impact**:Middle **Probability**:High Autodesk Revit has not a particular migration function. It can be therefore assumed that the exporting function might not be suitable for preservation actions anyhow.	This software should not be considered as possible tool for format migration. Rather another (independent) tool has to be developed and tested which enables a lossless migration from one IFC version into another IFC version.	High (Observed)
Since there are no suitable migration tools for IFC files in place, Autodesk Revit was the most feasible way to realize this sample preservation plan. Anyhow this investigation has shown, that the software is indeed not suitable for preservation actions.				

Table 1: Risks identified and assessed.

1.4 Technical Decisions

Rosetta as system for preservation planning assessment

As described in Deliverable D2.2.3[13], Rosetta is used within DURAARK as a DPS. Consequently, the preservation planning will also be conducted inside Rosetta. Rosetta provides tools, which support the selection of sample records and the decision making during the preservation planning process.

Autodesk Revit as tool for preservation action

As described in the previous section, 'Autodesk Revit' was used as a tool for the migration of IFC2x3 files into IFC4 files.

Besides, this work picked up on the following technical decisions made in the project:

- The work is based on the developed and evaluated function of the DURAARK WorkbenchUI which enables the generation and deposit of architectural 3D data containing SIPs.
- The RDF snapshots stem from the SDA as described in Deliverable D3.3[14].
- The technical metadata *ifcm*, like described in Deliverable D6.2, was used as technical input factor for the selection of significant properties.

The outcome of this work is integrated into the WP6 task 6.4, where a sample preservation planning for 3D objects was required.

[13] `http://duraark.eu/wp-content/uploads/2014/02/duraark_d2.2.3_final.pdf`
[14] `http://duraark.eu/wp-content/uploads/2014/08/DURAARK_D3_3_3.pdf`

2 Use cases

In Section 4 of Deliverable D2.2.1[15] nine use cases were described, addressing the creation and consumption of architectural 3D data. The use cases were established for defining the interactions of the DURAARK system between users of the system, the DPS and external knowledge bases for some cases. They help to understand the functions and processes which have been developed and established in the WorkbenchUI and provide an overview of the involved stakeholders, preconditions and expected results.

This deliverable addresses two core long term preservation use cases:

- UC1: Deposit 3D architectural objects
- UC3: Maintain Semantic Digital Archive

The following section describes how these two use cases have been considered and addressed in this deliverable respectively how they have affected aspects of the preservation planning of 3D objects.

2.1 UC1: Deposit 3D architectural objects

This use case is only partly addressed in this deliverable but should mentioned nevertheless. The whole sample preservation plan is built upon the assumption that architectural 3D data was ingested into the Rosetta system, as described in Deliverable D2.5 and evaluated in Deliverable D7.4. Hence, the data for the preservation planning process was already processed through the WorkbenchUI. Additionally, technical metadata was extracted within this pre-ingest stage, which is necessary for a successful assessment of the sample preservation plan.
However, within this sample preservation plan architectural 3D data will only be submitted for trial and not stored permanently.

[15] http://duraark.eu/wp-content/uploads/2014/02/DURAARK_D2.2.1.pdf

2.2 UC3: Maintain Semantic Digital Archive

Even though this deliverable mainly features a sample preservation plan for architectural 3D data, Section 5.3.2 provides policies for 'SDA / SDO management and interactions' and formulates an approach for a sustainable preservation of the developed knowledge base. With the help of these policies a proof of concept was developed for the preservation of RDF snapshots taken from the SDA which is described in Section 6.2. However, this is only a first approach with regards to long term preservation of semantically enriched content as it just addresses the bit preservation. But it enables the maintenance of the SDA. Outside the scope of DURAARK it is recommended to increase the investigation towards the goal of logical and preservation, for instance to elaborate on possible, sustainable file formats for RDF snapshots which than can be recorded in PRONOM[16].

[16] https://www.nationalarchives.gov.uk/PRONOM/Default.aspx

3 Preservation Planning Process

As introduced in Deliverable D6.6.1 the Preservation Planning Process ensures the long term availability and usability of the preserved material. The preservation action itself is the result of an extensive and detailed Preservation Planing Process. Possible actions are

- **migration**, where a file format will be converted into another one,
- **emulation**, where a virtual software environment is created for the access of obsolete files or even
- **technical museums**, where not only the files are preserved but also hardware components and software that was used for their access and performance.

The last possibility is not a generic approach but in some cases it could be feasible to preserve the original environment for access over the long term. For the most part, migration or emulation are the preferred alternatives for preservation actions. In fact, migration is already in place in numerous software products, e. g. via plugins or export possibilities. Nevertheless, possibilities for emulation have increased in the last few years, too. The University of Freiburg has developed the architecture 'Emulation as a Service' which enables different emulation possibilities for the user.[17] But how to decide whether a migration or an emulation will meet the requirements for a sustainable preservation process? And how to identify possible risks or sources of errors?
At this point it becomes clear that a single action is not sufficient for digital preservation. Even if the appearance of the migrated object and its extracted metadata is unaffected by the preservation action since this can just be the first impression. Without a solid base of knowledge, significant information may be lost or more suitable alternatives are missed. In a worst case scenario, an inaccurately prepared decision may cause damage for the preserved material or even its loss. It is therefore of high importance, that every decision during the preservation process is explained, evaluated and possibly revised or rejected.

[17] `http://bw-fla.uni-freiburg.de/`

3.1 Prerequisites

Before coming to preservation planning and its processes, a number of prerequisites are identified and described in the following sections.
First, there is the 'Preservation staff', without who the required processes and decisions are not possible. Second there watch activities are described, named 'Technology Watch' and 'Community Watch' which make preservation planing activities and resulting actions possible in time.

3.1.1 Preservation staff

Before talking about preservation planning it is worth to emphasize the necessity of a well-trained staff which is able to identify the requirements for specific formats, the need for preservation actions due to the results of the 'Watch functionalities' and to maintain preparation, performance and the evaluation of all preservation planning activities. Additionally, the staff is responsible for the decisions whether a preservation action was successful and shall be finally executed. The management of a digital archive (or an institution which hosts a digital archive) is responsible for the selection, training and establishment of sufficient staff members. The importance of the preservation staff will be described in more detail in Section 4.2.3. In the following description of the DURAARK preservation workflow the staff will be referred to as 'staff' or 'evaluator'.

3.1.2 Watch activities

As stated in the OAIS reference model, an institution has to provide, watch or monitor functionalities which ensure the understandability of the preserved material over the long term even if the technology has changed or the requirements and knowledge base on the user's site.[5]

Community Watch
The users or the community - referred to as as 'Designated Community' in the OAIS - with their requirements for software and data are certainly the main indicator whether there is a risk that the preserved materials become unusable. These requirements have to be considered even if a specific file format or a specific processing software is well

accepted. If this is the case, the digital archive is prompted to evaluated alternatives.[13] For instance, it can be possible that no open source software is supported within an institution due to legal reasons and the most suitable file format is therefore probably not usable from a technical point of view.
Thus, is is necessary to define the community of the digital archive and get in touch with them for defining their specific requirements. This should be conducted frequently as the community might change over time with new stakeholders, disappearing stakeholders or even with completely revised or new technical requirements.[13] This can be realized with the help of direct feedback in the form of questionnaires as done for the Deliverables D7.3 or D7.4 where DURAARK stakeholders have filled questionnaires regarding long term preservation related questions.

Technology Watch
For preservation purposes, technology itself should be also be monitored. This task is referred to as 'Monitor Technology' in the OAIS . Within this task, an institution should provide a service which helps to identify changes of technologies, file formats or standards.[13]

Here, 'Changes' refer for instance to the following:

- a new version of a file format is released
- previously used software becomes obsolete and is replaced by another one or a higher version
- a new standard for a specific file format is released
- a preserved file format is not supported anymore

3.2 Plato Workflow

As already described in Deliverable D6.6.1 the preservation planning tool 'Plato' was a major outcome of the PLANETS project and was further developed within SCAPE. Its state as an outcome of two European projects underlines its importance and the knowledge already invested. With 'Plato', a first approach for the decision process and methodology[18] during the preservation planning was done, which is why this tool can be referred to as 'State of the Art' when it comes to preservation planning. Hence, it is consistent to take the developed workflow as base for preservation planning in the DURAARK project.

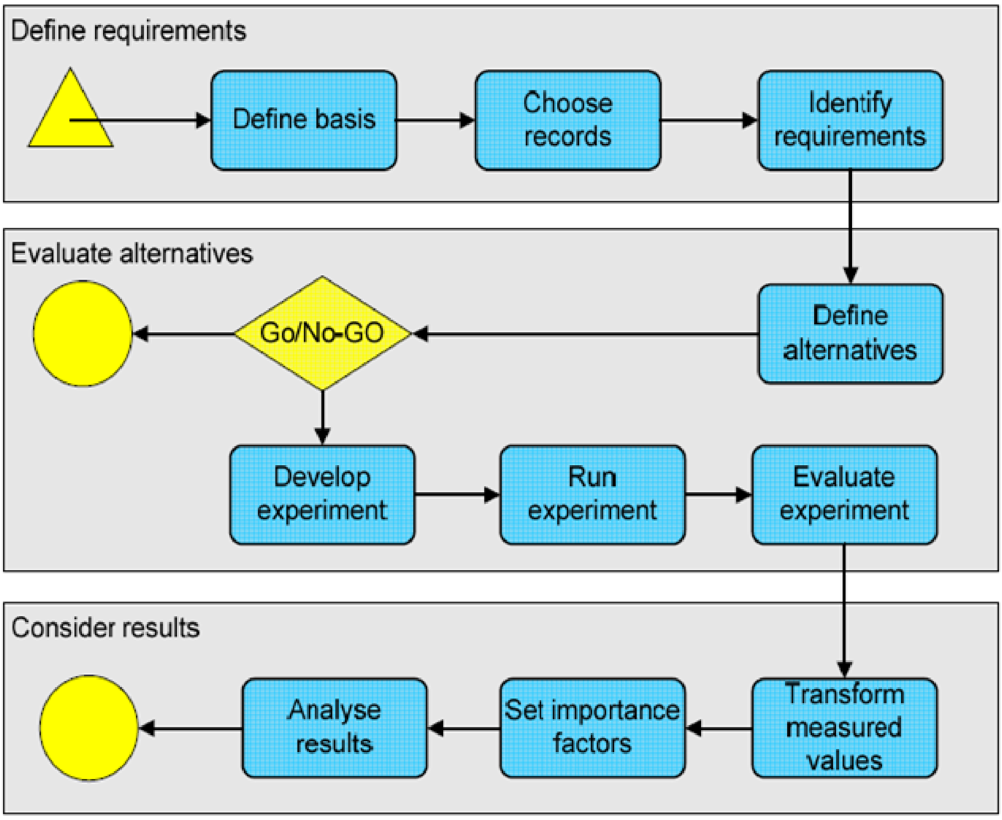


Figure 1: Preservation planning with Plato

'Plato' is a decision support tool which supports testing and evaluation through the whole preservation planning process. The whole workflow of Plato is illustrated in Figure 1 and is divided in three high level stages:

1. **Define requirements:** All information of importance for preservation of digital material should be gathered. This includes technical requirements, but also legal constraints and policies. Of course, the data for the evaluation should be in place as well.[3]

[18]Initial work was conducted within the DELOS project which was a starting point for the work in PLANETS as well.

2. **Evaluate alternatives:** The potential strategies and implemented tools have to be conducted in a test environment. Only then the strategy can be evaluated against the defined requirements.[3]

3. **Consider results:** Since some of the requirements may carry more weight than others, they have to be compared by the evaluator for making a final decision: Whether the evaluated preservation plan was successful and can be performed or if the the results show inconsistencies and possible errors which may lead into information loss.[3]

3.3 DURAARK Preservation Workflow

As stated in the previous section, DURAARK has decided to reuse the methodology which was provided by the 'Plato' project. But instead of using the decision tool directly, DURAARK suggests a workflow that is based on the 'Plato' development. The following section offers a generic approach for the preservation planning assessment for DURAARK purposes. Since DURAARK uses Rosetta as DPS, the presented workflow mentions it as a supporting tool as well. However, other systems or tools can potentially be used for the assessment during preservation planning. Such alternatives are however out of scope of the DURAARK Preservation Workflow.

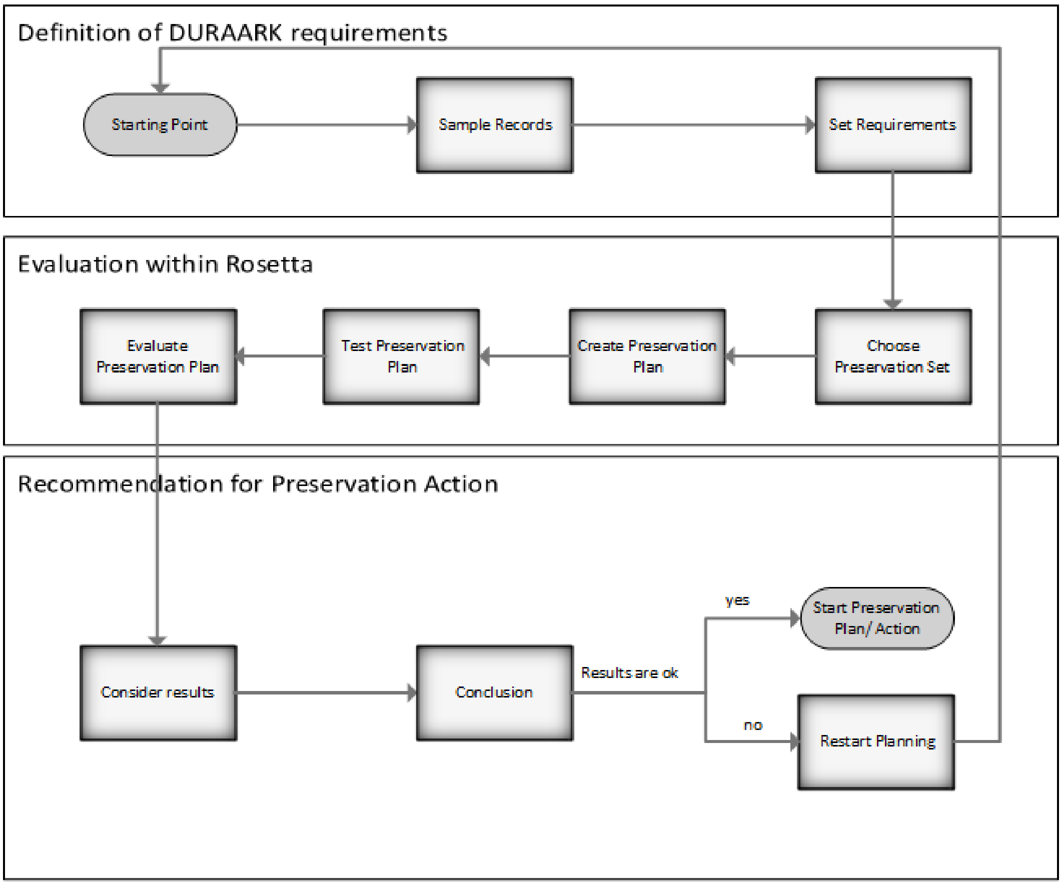


Figure 2: Preservation Planning in DURAARK

As shown in Figure 2 the structure looks slightly different. In the following, it is described how DURAARK has adopted and simplified the Plato workflow. DURAARK does not use an automatized tool for assessment purposes. Instead, decisions are made manually with the help of the Rosetta system and technical metadata extractors.

3.3.1 Definition of DURAARK requirements

At first, the requirements for a specific preservation plan have to be specified. This applies to both Plato and the DURAARK workflows.

- **Starting point**
  The Starting point in the DURAARK workflow are a set of a precise requirements for the evaluation and performance of a preservation plan. The focus of the DU-

RAARK project is set on the following data formats, as described in detail in the Deliverable D2.1:

- IFC for Building Information Models and
- E57 for Point Cloud Scans.[19]

In any case it is important to motivate the need for a preservation plan. As described in Section 3.1.2 within the OAIS this is considered as 'Monitor Designated Community' or 'Monitor Technology' for monitoring the environment of an institution respectively of a digital archive.[5] For instance, the obsolescence of a preserved file format is a motivation for a preservation planning process. A previously failed preservation plan may also be a motivation for a new preservation planning process.

- **Sample records**
  Based on the previously mentioned starting point, sample records[20] in the digital archive should be gathered and provided for further planning - the more, the better. These are not necessarily the records, used during the test preservation, but some, which provide technical characteristics for the next step.

- **Set requirements**
  Within Plato the *"requirements definition is the heart of preservation planning."*[2] And so it is in the DURAARK project.

[19]For this deliverable only a planning process for IFC files will be conducted.

[20]Sample records are already part of the digital archive and are used within the preservation planning process for test purposes. The requirements, which have been defined as part of the starting point (e.g. a file format is obsolete; a file format is not supported anymore) apply on these records.

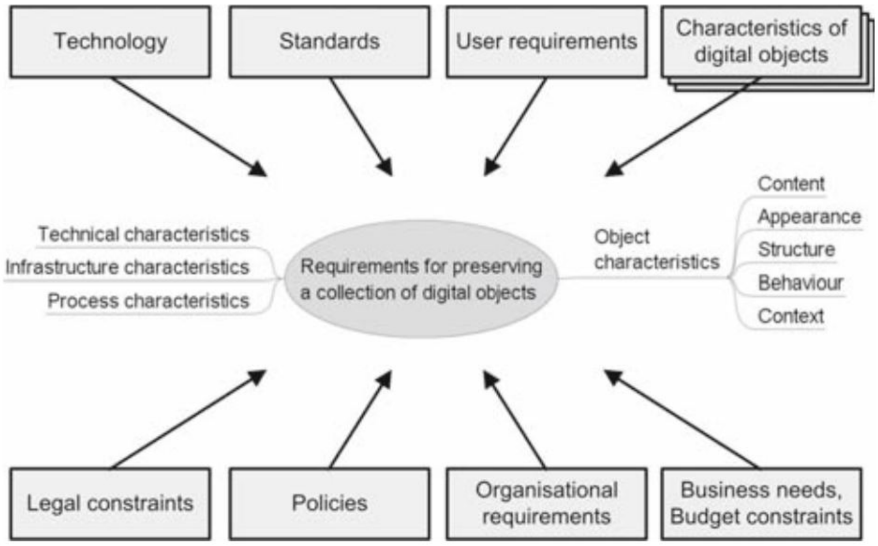


Figure 3: Influence factors (after Plato)[2]

As shown in Figure 3, also taken from Plato, several influences might have an impact on the requirements for preservation planning such as for instance 'technology', 'characteristics of digital objects' and 'policies'. They have to be collected and rated for having an overview, and have an outstanding influence on the success of a preservation plan.

For the sample preservation plan of 3D objects, DURAARK focuses on significant properties and on policies as possible influences as they are the most important and often referenced input factors for requirements. Event though other factors and constraints might affect the preservation planning, the advantage of policies - respectively their containing Control Policies which will described further in Section 5.2 - and significant properties is their measurability.[17] [13] By focusing on significant properties as input factors, the project considers not only the technical characteristics of the objects, but also organizational factors and requirements of stakeholders. The defined requirements are explained in detail and presented in Section 4 and Section 5.

3.3.2 Evaluation within Rosetta

Within Rosetta the whole sample preservation action will be conducted with the help of the system and its tools. Rosetta offers some functionalities for identifying risks on objects or comparing data before and after a preservation action.

- **Chose preservation set**
  Within Rosetta, sample records have to be selected again. Contrary to the second stage, this time records are not requested for the definition of requirements, but for the test run of a preservation plan. However, they should meet the needs for the scenario to be tested. Hence, it is a prerequisite that some relevant records are already present in the DPS.

- **Create preservation plan**
  Rosetta helps to build up a preservation plan based on the previously selected preservation set. Stakeholders can now select if the preservation plan will be run automatically or manually. Since the metadata extractors for IFC and E57 files developed within DURAARK are not yet in place in the Rosetta system[21], the assessment will be conducted manually. It is also possible to select whether a migration will be executed by plugins already implemented[22] or if it will be done externally in a specific software.

- **Test preservation plan**
  As a next step, the entire preservation plan will be performed as a test within Rosetta. If a migration has to be conducted externally, the evaluator is now prompted to do so and to import the newly created file into the Rosetta system.

- **Evaluate preservation plan**
  After the sample preservation plan has been executed internally or externally, it is possible to be evaluated with the help of Rosetta functions. If the test scenario was successful and is ready for a real preservation plan inside the system, it can be signed off, if not, it can be rejected. Since further evaluation takes place in the next stage, it can be left open until a final decision on the plan was made.

[21] The reasons why this is not done yet is described in detail in Deliverable D7.4.

[22] A partner of TIB Hannover uses for instance a plugin for creating valid PDF files. This is useful if the original files are not valid and therefore not sufficient for preservation over the long term.

3.3.3 Recommendation for preservation action

After the test-runs within Rosetta is finished, they have to be evaluated by the preservation staff with a conclusion whether the preservation action is applicable, should be revised or even rejected.

- **Consider results**
  All gathered information and results shall be extensively evaluated against the previously defined requirements. As outlined in Section 3.3.1, the policies and significant properties are partly measurable. Therefore, an easy and fast assessment is possible. Dissimilarities can easily be identified and make the evaluation of the preservation action itself more comfortable. If for instance, one defined significant property has another value than before the action, the staff is prompted to investigate this issue. The same applies for policies that are not fulfilled. Failures may occur due to different reasons - potentially the preservation action itself doesn't fit the needs or the used tools for assessment. It is therefore necessary to identify the cause of these negative results for leading into a sustainable decision for the next step and the preserved materials.
- **Conclusion**
  Based on the evaluated results, the staff shall decide, whether the preservation plan can be executed as-is or if a new preservation plan has to be built up. The latter is possible, if for instance some defined significant properties have been evaluated with a negative outcome. However, this decision shall be well and sustainable documented within the Rosetta system.

3.4 Conclusion

Even though the provided workflow is based upon the work which was conducted for Plato it takes the DURAARK environment into consideration. Additionally, it forgoes additional technical infrastructure which is required for the performance of Plato. This workflow was used for the sample preservation planning which is described in Section 6.1.

4 Significant Properties

Significant properties have been described in Section 2.6.2 of Deliverable D6.6.1 as measurable characteristics of an object, environment or even characteristics of the preservation action itself, which have to be preserved over time. Since the chosen characteristics are based on the needs and requirements of one or more stakeholders, they are often subjective and generic only sometimes.
Significant properties are often considered as technical characteristics which have to be preserved over time through any preservation action which might occur. Characteristics of an object or technical decision which has to be made, have direct or indirect impact on the organization itself. This means that significant properties have to be considered as technical as well as organizational input factors.
The following section provides an explanation on significant properties as technical and organizational input factors with regards to the sample preservation plan of IFC files which is presented in Section 6.1.

4.1 Technical input factors

Object-based characteristics mostly stem from the technical metadata which is extracted from the digital object itself but can also include descriptive metadata, for instance a file name which describes the content of an object like '2015-04-05-Scan-of-Cologne-Cathedral.e57'. In an ideal world, every property of an object should be preserved over the long term to enable its sustainable digital preservation. Unfortunately, this is not possible as it requires personal knowledge, financial resources as well as the technology for extracting, performing or storing all the properties which stem from an object. The more complex the object's composition is, the more difficult it is to preserve all of its properties. Therefore it is necessary to select the properties with which the authentic presentation and behaviour of the digital object is secured also after several preservation actions. 'Significance' is relative and subjective and hence, a universal and unchanging selection of significant properties is not feasible. Thus, it is necessary to evaluate these properties carefully and re-evaluate them during the whole preservation process.
InSPECT investigated the relevance of significant properties for digital preservation, dealing with its determination and measurement for specific file groups. To this end, the

project has chosen audio, email, raster images and structured text objects for investigation. InSPECT focused on the technical 'characteristics of digital objects that must be preserved over time in order to ensure the continued accessibility, usability, and meaning of the objects, and their capacity to be accepted as evidence of what they purport to record.'[10]

Defined properties stem from the technical information of the digital objects. For instance, the following properties were defined for raster images: Image Width, Image Height, X Sampling Frequency, Y Sampling Frequency, Bits per sample, Sample per pixel and Extra sample.[10] Instead of investigating properties of particular raster images the project covered the highest level of all analysed raster images.
However, this is only a subjective recommendation as in every institution different constraints or requirements exist, which may cause different selections for significant properties. InSPECT also provides two workflows for supporting the definition of significant properties:

- one workflow for the analysis of digital objects and
- one workflow for the analysis of the stakeholder requirements.

The workflow for the analysis of digital objects was used as methodology for the identification of significant properties within the DURAARK project. The workflow for the analysis of stakeholder requirements is further described in Section 4.3.

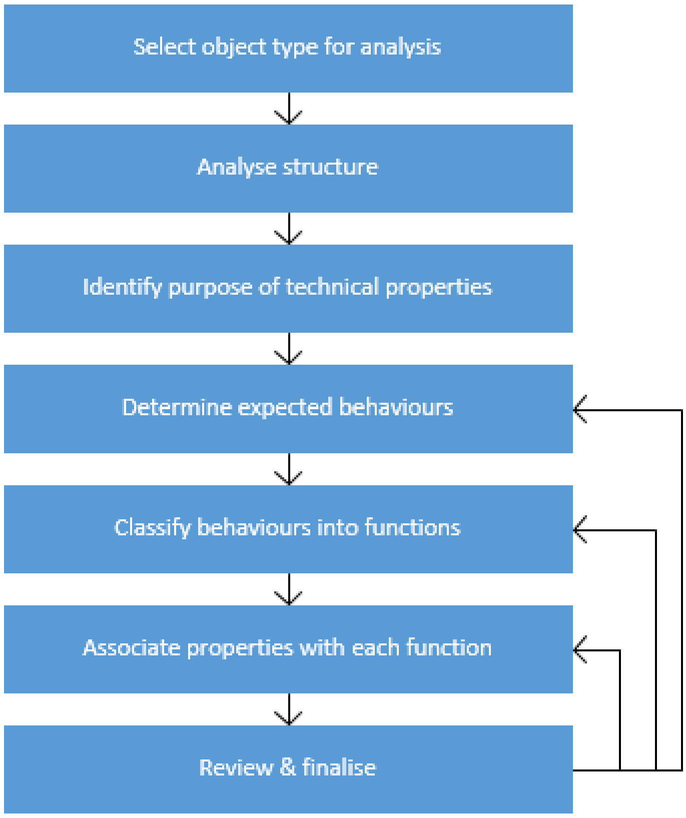


Figure 4: Workflow for the object analysis based on InSPECT[11]

The workflow for object analysis supports the evaluator for the selection of significant properties. As shown in Figure 4 it consists of seven tasks which may be executed sequentially.

In the *first step* the object type which should be analysed has to be defined. This could be high level - for instance raster images, building information models - or more specific - for instance JPEG or IFC.

Within the *second step* the evaluator has to analyse the structure and composition of the file format. InSPECT recommend the use of additional tools for extracting and providing technical metadata.

In the *third step* the staff should investigate and identify the purpose and usage of each technical property. The project defined five high level categories 'content', 'context', 'rendering', 'structure' and 'behaviour' for a distinct analysis and identification of the different specific roles an extracted information may play within a digital object.

Within the *fourth step* the evaluator considers the different kind of activities any stake-

holder may want to perform with the file format to generate a check list of possible behaviours of the digital object.
These identified behaviours shall be classified as functions within the *fifth step*. InSPECT mention for instance the 'Visual presentability' or 'Authenticity verification' as functions. These functions are mostly specific and dependent of the investigated object.
The *sixth step* interlinks the already identified technical properties with a set of expected behaviours. By preserving these interlinked properties, the evaluator ensures appropriate behaviour expected by by the stakeholder in the future.
In the *seventh step* the evaluator shall review and finalize the gathered information and check, if some behaviours have been overseen or some functions may be refined into more accurate ones. For instance the function 'Authenticity verification' can be devised in the function 'Integrity check' and 'Provenance verification.[11]

The workflow provided by InSPECT helps to define significant properties which stem from the digital object itself. For DURAARK's purpose it is necessary to define these properties starting with the object analysis.

4.1.1 Select object type for analysis

Since this deliverable provides a sample preservation plan for IFC files in particular, it focuses on the file format IFC and its corresponding versions 2x2, 2x3 and 4.

4.1.2 Analyse structure

A good starting point for analysing the structure is the technical metadata for IFC files, which was introduced in Deliverable D6.2 as *ifcm*.

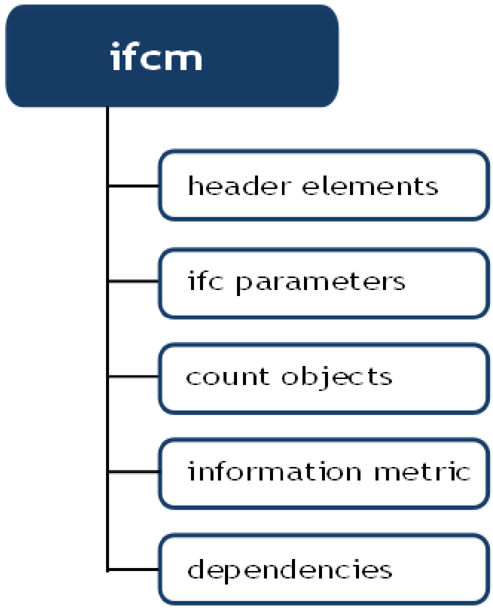


Figure 5: High level overview of the *ifcm* structure.

In Figure 5 a high level overview of the schema is provided. While the header 'elements' and 'ifc parameters' are extracted directly from the IFC elements, the other elements like for instance count objects'[23] are processed from the IFC elements. Hence the following elements of *ifcm* can be considered as potentially significant properties:

- ifcm:header:
  - name, creationDate, author, organization, preprocessor, originatingSystem, authorization, fileSchema, viewDefinition, exportOptions
- ifcm:ifcparameters:
  - IfcApplication, IfcGeometricRepresentationContext, IfcSiUnit
- ifcm:countObjects:

[23]In the latest version of *ifcm* this part was renamed into 'quantities' to be found under following URL: `https://github.com/DURAARK/Schemas/blob/master/xml/2015_03_30_duraark_ifcm_1_1.xsd`.

  - floorCount, roomCount, wallCount. windowsCount, doorCount, pipeCount, columnCount, numberOfComponents, numberOfRelations, numberOfActors

- ifcm:InformationMetric:
  - numberOfEntityTypesUsed, numberOfTotalEntitiesUsed, optionalAttributes

- ifcm:Dependencies:
  - webResourceLink

Beside the elements which are presented as *ifcm*, there are other high level file properties which have to be considered as potential significant properties:

- file name
- file extension
- file size
- format ID (taken from PRONOM)
- file fixity (can be MD5, SHA1 or CRC32)

4.1.3 Identify purpose of technical properties

As described above, InSPECT recommends to categorize the identified elements for distinguishing into 'content', 'context', 'rendering', 'structure' and 'behaviour'. For the DURAARK context, this was adapted accordingly and is provided in Table 2:

class	**technical property**
content	• ifcm:ifcparameters:IfcGeometricRepresentation-Context • ifcm:ifcparameters:IfcSiUnit • ifcm:countObjects:floorCount • ifcm:countObjects:roomCount • ifcm:countObjects:wallCount • ifcm:countObjects:windowsCount • ifcm:countObjects:doorCount • ifcm:countObjects:pipeCount • ifcm:countObjects:columnCount • ifcm:countObjects:numberOfComponents • ifcm:countObjects:numberOfRelations • ifcm:countObjects:numberOfActors • file size • file extension • format ID (taken from PRONOM)

context	• ifcm:header:name • ifcm:header:creationDate • ifcm:header:author • ifcm:header:organization • ifcm:header:authorization • file name
rendering	• ifcm:header:preprocessor • ifcm:header:originatingSystem • ifcm:header:fileSchema • ifcm:header:viewDefinition • ifcm:header:exportOptions • ifcm:ifcparameters:ifcApplication
structure	• ifcm:InformationMetric:numberOfEntity TypesUsed • ifcm:InformationMetric:numberOfTotal EntitiesUsed • ifcm:InformationMetric:optionalAttributes • file fixity (MD5, SHA1 or CRC32)
behaviour	• ifcm:Dependencies:webResourceLink

Table 2: InSPECT object categorization

4.1.4 Determine expected behaviours

Several behaviours can be associated with the IFC format. Here, Behaviours mean generic interactions that are possible with the file. The following behaviours have been identified:

- Save the IFC file on a storage medium.
- Exchange the IFC file with other potential stakeholders.
- Reproduce the visual appearance of the IFC file with the help of software as originally intended.
- Edit the IFC file with the help of software.
- Identify the detailed file format and version of the IFC object with the help of a format identification tool.
- Access the IFC file with the help of a text editor.
- Create new IFC files based on one existing IFC file.
- Extract technical metadata from the IFC file.
- Link to external web resources which are placed inside the IFC file.
- Import the IFC file in a CAD software.
- Reproduce the provenance trail of the IFC file with its creation, every manipulation, migration, addition etc.
- Create checksums of the IFC file.
- Keep the file unchanged.

4.1.5 Classify behaviours into functions

For ensuring its future usage, InSPECT has recommended to classify the identified behaviours into functions. A single behaviour can be classified as more than just one function. In the second column of the following table 3 functions are formulated which are based on the behaviours that have been identified in the previous section.
The functions - like for instance 'Verify authenticity' - are strongly connected to the identified significant properties. This will be described in detail in Section 4.1.6.

function	**behaviours**
Verify authenticity	• Reproduce the visual appearance of the IFC file with the help of software as originally intended. • Create checksums of the IFC file. • Reproduce the provenance trail of the IFC file.
Verify integrity	• Create checksums of the IFC file. • Save the IFC file on a storage medium. • Keep the file unchanged.
Rendering	• Access the IFC file with the help of a text editor. • Reproduce the visual appearance of the IFC file with the help of software as originally intended. • Import the IFC file in a CAD software.
Exchange	• Exchange the IFC file with other potential stakeholders. • Import the IFC file in a CAD software.
Editing	• Edit the IFC file with the help of software. • Create new IFC files based on one existing IFC file.

Characterization	• Identify the detailed file format and version of the IFC object with the help of a format identification tool. • Extract technical metadata from the IFC file.
Understanding	• Reproduce the visual appearance of the IFC file with the help of software as originally intended. • Access the IFC file with the help of a text editor. • Link to external web resources which are placed inside the IFC file. • Reproduce the provenance trail of the IFC file with its creation, every manipulation, migration, addition etc.

Table 3: InSPECT behaviour categorization

4.1.6 Associate structure with each behaviour

Long term preservation aims at life-cycles of hundreds of years. Therefore, the authenticity of a preserved IFC file is hard to verify only with the help of the visual appearance of an IFC file in 100 years. A comparison with its appearance nowadays is nearly impossible, because the software for access will be for sure another one or at least available in another version. Hence, it is important to save properties along the file which guarantee the visual appearance of the file over time. These properties will be extracted before and after each manipulation on the file, which might be the case due to preservation actions. As a result the classified structure of the IFC file - which was presented in Section 4.1.2 and Section 4.1.3 - are finally associated with the identified functions and behaviours - which were presented in the previous Section 4.1.5 - to have a summery of which properties are required for which function. This is presented within the following tables. Due to layout reasons every function is features in a single table with its associated structure.

Verify authenticity		
behaviours	**property class**	**properties**
Reproduce the visual appearance of the IFC file with the help of software as originally intended.	Content	ifcm:countObjects:floorCount ifcm:countObjects:roomCount ifcm:countObjects:wallCount ifcm:countObjects:windowsCount ifcm:countObjects:doorCount ifcm:countObjects:pipeCount ifcm:countObjects:columnCount ifcm:countObjects:numberOfComponents ifcm:countObjects:numberOfRelations ifcm:countObjects:numberOfActors file size file extension format ID (taken from PRONOM)
	Structure	ifcm:InformationMetric:numberOfEntityTypesUsed ifcm:InformationMetric:numberOfTotalEntitiesUsed ifcm:InformationMetric:optionalAttributes
Create checksums of the IFC file.	Structure	file fixity (MD5, SHA1 or CRC32)
Reproduce the provenance trail of the IFC file.	Context	ifcm:header:name ifcm:header:creationDate ifcm:header:author ifcm:header:organization ifcm:header:authorization file name

Table 4: Associated structure with function 'Verify authenticity'

Verify integrity		
behaviours	**property class**	**properties**
Create checksums of the IFC file.	Structure	file fixity (MD5, SHA1 or CRC32)
Save the IFC file on a storage medium.	Content	file extension file size
Keep the file unchanged.	Structure	file fixity (MD5, SHA1 or CRC32)
	Content	file extension file size format ID (taken from PRONOM)
	Context	ifcm:header:name ifcm:header:creationDate
	Structure	ifcm:InformationMetric:numberOfEntityTypesUsed ifcm:InformationMetric:numberOfTotalEntitiesUsed ifcm:InformationMetric:optionalAttributes

Table 5: Associated structure with function 'Verify integrity'

Rendering		
behaviours	**property class**	**properties**
Access the IFC file with the help of a text editor.	Content	file extension file size format ID (taken from PRONOM)

Reproduce the visual appearance of the IFC file with the help of software as originally intended.	Content	ifcm:countObjects:floorCount ifcm:countObjects:roomCount ifcm:countObjects:wallCount ifcm:countObjects:windowsCount ifcm:countObjects:doorCount ifcm:countObjects:pipeCount ifcm:countObjects:columnCount ifcm:countObjects:numberOfComponents ifcm:countObjects:numberOfRelations ifcm:countObjects:numberOfActors file size file extension format ID (taken from PRONOM)
	Structure	ifcm:InformationMetric:numberOfEntityTypesUsed ifcm:InformationMetric:numberOfTotalEntitiesUsed ifcm:InformationMetric:optionalAttributes
	Rendering	ifcm:header:preprocessor ifcm:header:originatingSystem ifcm:header:fileSchema ifcm:header:viewDefinition ifcm:header:exportOptions ifcm:ifcparameters:ifcApplication
Import the IFC file in a CAD software.	Rendering	ifcm:header:preprocessor ifcm:header:originatingSystem ifcm:header:fileSchema ifcm:header:viewDefinition ifcm:header:exportOptions ifcm:ifcparameters:ifcApplication

Table 6: Associated structure with function 'Rendering'

Exchange		
behaviours	**property class**	**properties**
Exchange the IFC file with other potential stakeholders.	Context	ifcm:header:name ifcm:header:creationDate ifcm:header:author ifcm:header:organization ifcm:header:authorization
	Content	file extension format ID (taken from PRONOM) file name
Import the IFC file in a CAD software.	Rendering	ifcm:header:preprocessor ifcm:header:originatingSystem ifcm:header:fileSchema ifcm:header:viewDefinition ifcm:header:exportOptions ifcm:ifcparameters:ifcApplication
	Content	file extension format ID (taken from PRONOM)

Table 7: Associated structure with function 'Exchange'

Editing		
behaviours	**property class**	**properties**
Edit the IFC file with the help of software.	Rendering	ifcm:header:preprocessor ifcm:header:originatingSystem ifcm:header:fileSchema ifcm:header:viewDefinition ifcm:header:exportOptions ifcm:ifcparameters:ifcApplication
	Content	file extension format ID (taken from PRONOM)

Create new IFC files based on one existing IFC file.	Rendering	ifcm:header:preprocessor ifcm:header:originatingSystem ifcm:header:fileSchema ifcm:header:viewDefinition ifcm:header:exportOptions ifcm:ifcparameters:ifcApplication
	Content	file extension format ID (taken from PRONOM)

Table 8: Associated structure with function 'Editing'

Characterization		
behaviours	**property class**	**properties**
Identify the detailed file format and version of the IFC object with the help of a format identification tool.	Content	file extension format ID (taken from PRONOM)
Extract technical metadata from the IFC file.	Content	file extension format ID (taken from PRONOM)

Table 9: Associated structure with function 'Characterization'

Understanding		
behaviours	**property class**	**properties**
Reproduce the visual appearance of the IFC file with the help of software as originally intended.	Content	ifcm:countObjects:floorCount ifcm:countObjects:roomCount ifcm:countObjects:wallCount ifcm:countObjects:windowsCount ifcm:countObjects:doorCount ifcm:countObjects:pipeCount ifcm:countObjects:columnCount ifcm:countObjects:numberOfComponents ifcm:countObjects:numberOfRelations ifcm:countObjects:numberOfActors file size file extension format ID (taken from PRONOM)
	Structure	ifcm:InformationMetric:numberOfEntityTypesUsed ifcm:InformationMetric:numberOfTotalEntitiesUsed ifcm:InformationMetric:optionalAttributes
	Rendering	ifcm:header:preprocessor ifcm:header:originatingSystem ifcm:header:fileSchema ifcm:header:viewDefinition ifcm:header:exportOptions ifcm:ifcparameters:ifcApplication
Access the IFC file with the help of a text editor.	Content	file extension file size format ID (taken from PRONOM)
Link to external web resources which are placed inside the IFC file.	Behaviour	ifcm:Dependencies:webResourceLink

Reproduce the provenance trail of the IFC file.	Context	ifcm:header:name ifcm:header:creationDate ifcm:header:author ifcm:header:organization ifcm:header:authorization file name

Table 10: Associated structure with function 'Understanding'

4.1.7 Review and finalise

Finally, the identified properties have been cross matched with the identified functions and behaviours. If for instance the community expects 'Authenticity' or 'Rendering' as a provided function of the preserved objects, the matrices presented in the previous sections can be used as a basis for deriving the properties which have to be preserved.
Since not every ifcm element is a mandatory one it is hard to define the definitive significant properties. As discussed above, nearly every extracted property is important for most of the identified functions. For DURAARK's purpose it is important to save the integrity and authenticity of the preserved material. Additionally, an IFC file has to be rendered with suitable software and keeps its visual appearance as originally intended. Hence, it is strongly recommended to preserve and check at least all technical properties during the whole preservation process which have been cross matched in Table 4, Table 5 and Table 6:

- ifcm:countObjects:columnCount
- ifcm:countObjects:doorCount
- ifcm:countObjects:floorCount
- ifcm:countObjects:numberOfActors
- ifcm:countObjects:numberOfComponents
- ifcm:countObjects:numberOfRelations
- ifcm:countObjects:pipeCount
- ifcm:countObjects:roomCount
- ifcm:countObjects:wallCount

- ifcm:countObjects:windowsCount
- ifcm:header:author
- ifcm:header:authorization
- ifcm:header:creationDate
- ifcm:header:exportOptions
- ifcm:header:fileSchema
- ifcm:header:name
- ifcm:header:organization
- ifcm:header:originatingSystem
- ifcm:header:preprocessor
- ifcm:header:viewDefinition
- ifcm:ifcparameters:ifcApplication
- ifcm:InformationMetric:numberOfEntityTypesUsed
- ifcm:InformationMetric:numberOfTotalEntitiesUsed
- ifcm:InformationMetric:optionalAttributes
- file extension
- file fixity (MD5, SHA1 or CRC32)
- file name
- file size
- format ID (taken from PRONOM)

The solutions presented here are specific to the DURAARK requirements may be completely different depending on whether a stakeholder focuses more on just the bit preservation of the IFC files which is why the presented significant properties must not be considered as a generic set for preservation planning of IFC files.
Within test-runs of the preservation process and its results it may occur that some of these properties and the corresponding functions cannot be preserved through the preservation action. In this case it is recommended to evaluate alternative preservation actions and run the test again. If it becomes clear that some of the properties cannot be preserved during the preservation action it is necessary to re-evaluate the technical input factors and consider - in agreement with stakeholders - if some of the impact factors

should be weighted stronger as others. If for instance the *authenticity* of the object is more important for the stakeholders than the *accessibility*, the access-related properties are of less significance.
This evaluation, re-evaluation an weighting is highly specific and dependent of several factors like the preservation planning results, the preservation staff, the stakeholders and of course the objects. Hence, it is not possible to give a general recommendation except for the necessity of a careful decision making process.

4.2 Organizational input factors

In the previous section, significant properties have been described as technical input factors and they are often considered as mere technical metadata which stem directly from the file format itself. Often it is overseen that some of the defined significant properties may have a direct or indirect impact on the organization itself and have therefore to be considered as organizational input factors.
In the following sections, some examples of organizational input factors are gathered which have to be considered when in comes to preservation planning and the preservation of significant properties. The presented factors are strongly subjective and not definitive and thus may overlap with each other here and there.

4.2.1 Finances and investment

Even in the OAIS it was stated, that the management of an archive has to be provided with budget for the development and implementation of standards and policies.[5] Hence, finances and investment are crucial input factors when it comes to digital preservation in general and to preservation planning in particular.
The implementation of a digital preservation itself requires lot of financial effort and the whole preservation planning process is cost-intensive as well. Since the preparation and performance of tasks like 'Community Watch', 'Technology Watch' or the development of preservation strategies take a well qualified staff for granted, the management also has to provide a budget for further training or specialized staff members. Certainly, this has not a direct, but an indirect impact on the selection for significant properties. According to InSPECT, an institution which already invested time, resources and money in building

up a digital archive and curating the digital material will also try to keep the preservation planning on this high level. Therefore they might intend to preserve as much properties as possible.[9]
Furthermore, the technical properties of a digital object may also have a direct impact on the finances or investments of the institution. Three dimensional data for instance is often considered as data objects with large file sizes and therefore require lot of storage space, especially if it is preserved in multiple versions. Hence, the file size itself must not be a direct significant property for the preservation of the object. But it is possible that a migration of a file into an equivalent or even bigger file format is taken into consideration as a preservation action. This may cause a bigger need of storage media, which is not only cost-intensive itself but also requires organizational input, like buying and installing the storage media. In consequence, the file size can be of significance for the institution which hosts the digital archive because of the investments in human resources and the media. Especially smaller institution are not able to provide this, which is why in the last years there was also discussion about storing data in the cloud.[7] Another side effect could be that the institution requires the migration in another file format with lower storage requirements.
As shown in Deliverable D7.3, Building Information Models are most frequently used *before* preservation actions are taken. Hence, the most budget may be invested in the 'productive' stage of the lifecycle rather than into the last stage. This may also have an impact on how to build up storage for preservation.

4.2.2 IT infrastructure

As mentioned above, a sustainable digital preservation and preservation planning is not possible without an appropriate IT infrastructure. This starts with the already mentioned storage media (e.g., mirrored hard disk drives, tape machines, optical jukeboxes etc.) to appropriate software for the curation and preservation of digital material up to special tools for particular preservation purposes like migration, identification and validation. Fortunately, numerous software solutions - like DROID[24], Archivematica[25] or JHOVE[26] - exist, which are Open Source and therefore do not require much financial efforts compared

[24] http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/
[25] https://www.archivematica.org/en/
[26] http://jhove.sourceforge.net/

to commercial solutions. Nevertheless, the infrastructure has to be in place to implement and provide the selected tools for preservation purpose.
As stated in Deliverable D6.6.1 the IFC SPFF files were validated by JHOVE as ASCII files due to their clear-text encoding. Also JHOVE tried to extract the metadata with its ASCII module - the extracted metadata was therefore only partly of interest for preservation purposes of IFC files. Hence, the selection of significant properties is strongly dependent of the tools that are available and in use, as they not extract every property which might be of significance for a particular object. Potentially, the organization feel obliged to develop or fund new tools for the extraction of any technical property which is needed for a sustainable digital preservation.[9]

4.2.3 Human resources

As already outlined in section 3.1.1 the preservation staff or human resources of an institution are one of the most crucial impact factors for the success or failure of all preservation related activities. As already written in the OAIS a digital archive it is not only technology, but the interaction between technology and the organization, the people which determines this success.[5] A digital archive causes a series of acts and decisions which have to be made by humans and which are not totally automatable. As described above, insufficiently qualified staff members may make false decisions. For instance, such a member selects the wrong types or wrong amount of properties which are of significance for preservation which finally lead into the damage or loss of the preserved material.
This again is not a significant property itself but influences the selection, preservation and future usage of the stored material.

4.3 Stakeholder requirements

As previously described, the presented input factors are very subjective and connected to the DURAARK view on IFC files respectively their preservation in a digital archive even though several, more objective factors where considered in the development of the use cases. However, there are several stakeholders, which may have a different view on the preserved material. Every stakeholder may have different needs, different interests, different requirements and different goals for preservation.

In the second year of the project, a questionnaire was developed within WP 6 and WP7 for the identification the needs of institutions who are working with architectural 3D data or at least plan to work with it soon. The results were used as a base for the Deliverable D7.3 and have given a first insight on how heterogeneous stakeholder requirements are.
In the third year of the project, again a questionnaire[27] was conducted for gathering a more detailed picture on how institutions work with their preserved materials and if there are policies respectively recommendations in place regarding digital preservation. The results were used for the evaluation activities in Deliverable D7.4 and have shown that 'Authenticity' and 'Integrity' was the most important aspect to be considered for the stored data for the majority of the stakeholders. However, one stakeholder considered 'Authenticity' as almost not important - the access and rendering on other hand to be of high importance.
The evaluator is encouraged to ensure that a set of significant properties will be preserved, which serves the needs and requirements of not only her or his organization, but also of any possible stakeholder, who might be interested in the preserved material. Since, this intention is hard to realize, the stakeholders have to be aware of the different target groups and their needs. This can be realized by stakeholder investigations, studying accessible materials of or on the stakeholders (e.g., description of goals, policies, legal information etc.) or establishing standards.

[27] https://docs.google.com/document/d/1P9Xi3YI7SRJBw_YV9Po7xH-NoNRe8PL85wk0w6KSjrQ

Again it was the InSPECT project, that has provided an approach for the identification the stakeholder's requirements on significant properties.

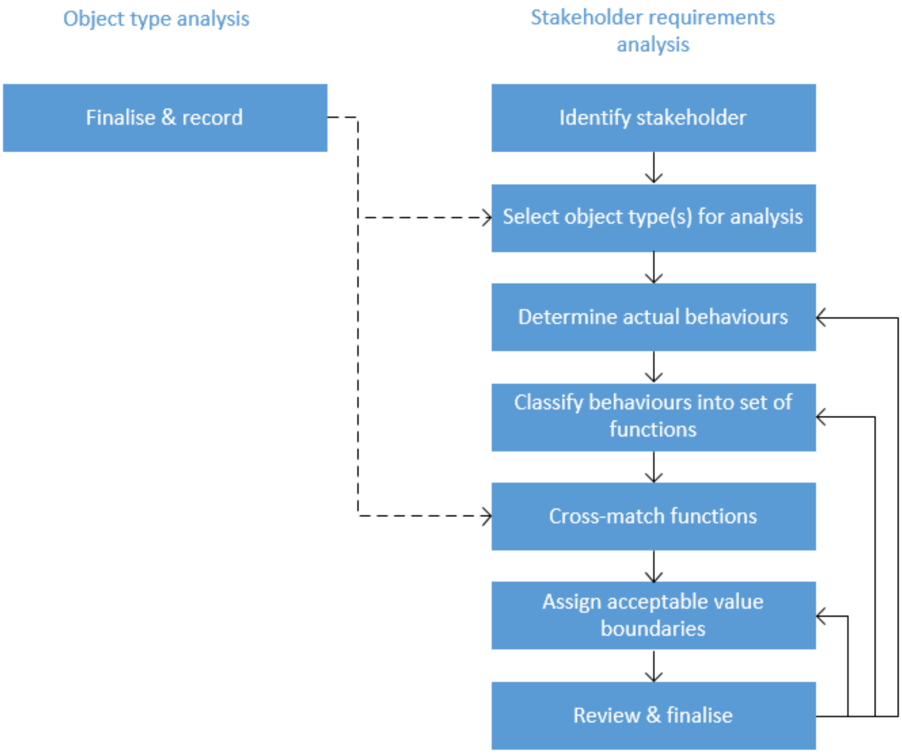


Figure 6: Work flow for the stakeholder requirements analysis based on InSPECT[11]

The provided work flow is shown in Figure 6 and consists of seven steps like the work flow in Figure 4.
In the first step the evaluator has to define the stakeholders which are targeted during the analysis. It is possible to have a broad variety of stakeholders since there are several stations, needs and requirements during the digital object's lifecycle. The evaluator may also consider existing policies or legal documents for this step.
As also done in the first step of the object analysis work flow, the evaluator has to select the object type for the stakeholder analysis. This can be based on the gatherings of the object analysis workflow.
In the third step the actual behaviours should be analysed and identified. In comparison to the step for expected behaviours of the object analysis workflow these behaviours are described as the 'real world behaviours'. Hence, a subset of the expected ones which is actually in use on the stakeholder's side.

In the next step the actual behaviours shall be classified as function, comparable to the object analysis work flow. InSPECT recommends to reuse the classification guide which was created as part of the object analysis work flow.
In the fifth step the technical properties shall be defined which are necessary for the performance of the functions required by the stakeholders. Hence, it is necessary as shown in the figure to compare the object functions with the stakeholder functions.[11]
This work flow can be used for the identification of stakeholder needs and the alignment with the previously identified object characteristics. However, the DURAARK project has already identified 'Authenticity', 'Integrity' and 'Rendering' in Section 4.1.7 as the most significant functions with its corresponding properties which have to be preserved. This decision was already made against the background of the stakeholder interviews which have been conducted for the work of Deliverable D7.3 and D7.4.

5 Preservation Policies

As stated in Section 3 and according to 'Plato', preservation policies are possible and necessary input factors for a successful digital preservation planning process. Beside the significant properties requirements are specified against which the results of a test preservation action are evaluated. This section discusses the use of preservation policies as input factors, presents an overview on the Preservation Policy Framework of the SCAPE project with its already existing generic policies for preservation and shows the DURAARK policies for preservation of 3D objects as well as for SDA / SDO management.

5.1 Policies as input factors

As already described in Section 4 in detail, significant properties are the most consulted input factors when discussing and measuring the success or failure of a preservation planning process. The goals and purpose of an institution for curating, storing and preserving digital materials and collections are important issues which cannot be neglected. Those purposes and goals are often captured with the help of (preservation) policies, created for the purpose of traceability. Thus, these documents shall be identified and considered as input factors for preservation planning.

5.1.1 Purpose

With the help of a preservation policy an institution ensures that the preservation of digital objects is explained and transparent for all involved stakeholders. Policies also help the institution to achieve its goals and establish its strategy as a cultural heritage institution. The documentation of theses facts is crucial since they constitute the awareness of the need of digital preservation itself.[20] They can be seen as a self-commitment of the institution with regards to digital preservation, with its consisting plans, workflows, strategies and actions. The institution's success is measured according to the policies in place.[19] [20] Furthermore, policies are useful for the preservation staff as an instrument for daily work.[18] They also help to evaluate preservation actions as they are partly measurable.

5.1.2 Usage

A preservation policy may support an institution by following a well-defined strategy and achieving its goals. Furthermore, policies have to be considered when it comes to preservation actions. Similar to legal constraints, all requirements and needs manifested in a policy have to be fulfilled.[18] Eventually, policies make the work and strategies of an institution also visible to a broader audience, like for instance the user of an archive. The user can decide, if the authenticity or integrity of the stored data is of high importance or if the institution has established functions and services for the long term availability of the preserved materials.[20]

5.1.3 Input factor

In Section 3.2 the SCAPE project was already mentioned. It was funded within the Seventh Framework Programme by the EU between 2010 and 2014 contributed among other things the 'Catalogue of preservation policy elements' (more on this in Section A.1) and has formulated a state of the art and overview when it comes to preservation policies. SCAPE characterises policies within the Preservation Watch of an institution as a necessary input factor, whose contained information is needed to *"formulate risks and constraints for the Preservation Planning activity."*[18]
The project 'Sustaining Heritage Access through Multivalent ArchiviNg' (SHAMAN) also funded within the Seventh Framework Programme, concluded, that preservation policies are *"essentially the instrument by which Governance Capabilities control Preservation Planning."*[1] The catalogue was built upon ideas which have been stated as an outcome of the PLANETS project. Hence, PLANETS also defined policies as a factor, which indicates the goals of an institution's digital preservation and therefore the constraints and requirements of a preservation planning process.[6]
It is generally agreed that preservation policies are a common, suitable and necessary input factor for digital preservation in general and the preservation planning in particular.[20] Since they are often depending on the preserved material and the specific goals of an institution, they are only partially generic and have to be adopted, re-written or even newly developed. However, several preservation policies are in place throughout Europe and are at least helpful for writing own institutional policies.

5.2 SCAPE Preservation Policy Framework

The 'Catalogue of preservation policy elements' also includes a description of the SCAPE Preservation Policy Framework which consists of three clustered policy levels for providing assistance in creating own preservation policies.[18] The DURAARK project follows the defined procedures here as it provides a highly structured and transparent approach and represents the current state of the art regarding preservation policies.

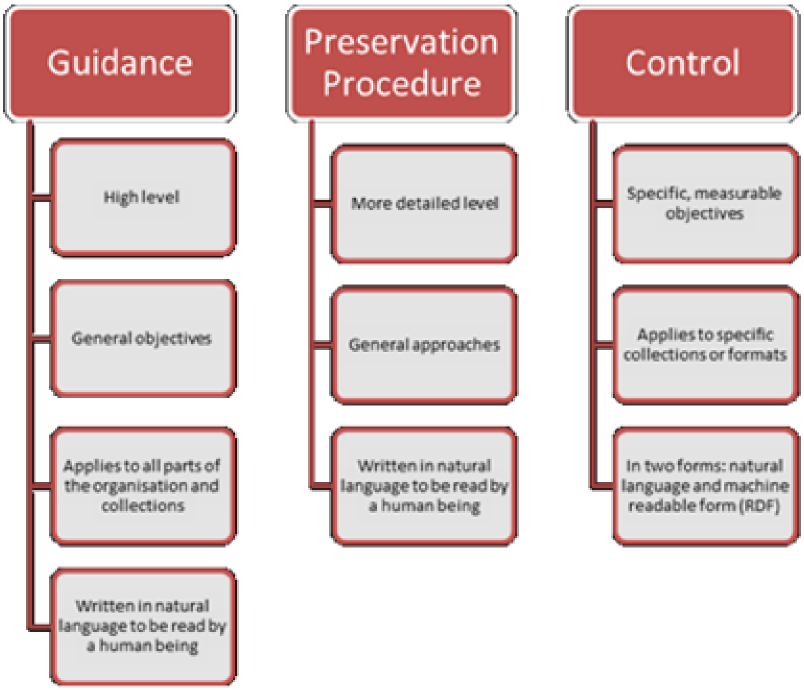


Figure 7: SCAPE Policy Framework

As shown in Figure 7 the three levels are named and described as:

- **Guidance Policy -** This level describes the general high level goals of an institution and its preservation approaches for the preserved material. SCAPE noted as an example, that an archive may be guided by the OAIS reference model.[18]
- **Preservation Procedure Policy -** For achieving the goals which, are aforementioned on the higher level, more specific and detailed approaches are formulated on this level in natural language.[18]
- **Control Policy -** The third level formulates the direct requirements for *"a specific*

collection, a specific preservation action or for a specific designated community." SCAPE stated that these policies can also be formulated in natural language but shall also be available as machine readable for enabling automation of planning and watch functionalities.[18]

SCAPE has developed and provided sample policies for other preservation aspects like 'Bit Preservation', 'Metadata' or 'Rights'. An overview to these policies can be found in Annex A.1. The provided policies also functioned as foundation for DURAARK's policies, which will be presented in the next section.

5.3 DURAARK Policies

For the DURAARK project, two Guidance Policies with containing Preservation Procedure Policies and Control Policies have been developed and created:

- **DURAARK Preservation:** This Guidance Policy describes with the help of its containing Preservation Procedure Policies the preservation related requirements and constraints which have to be considered with regards to digital preservation, preservation actions and therefore the whole preservation planning process.
- **SDA / SDO management and interactions:** This Guidance Policy are meant to specify the interaction between the SDA and the DPS.

DURAARK has used the 'policy element template'[28] which was also provided by SCAPE for a formalised description of the policy elements. The template can also be found in Annex A.2.
The template also recommends the usage of a lifecycle model for identify at which stage of the object's lifecycle a policy is relevant.[18]

[28] http://wiki.opf-labs.org/display/SP/Policy+element+template

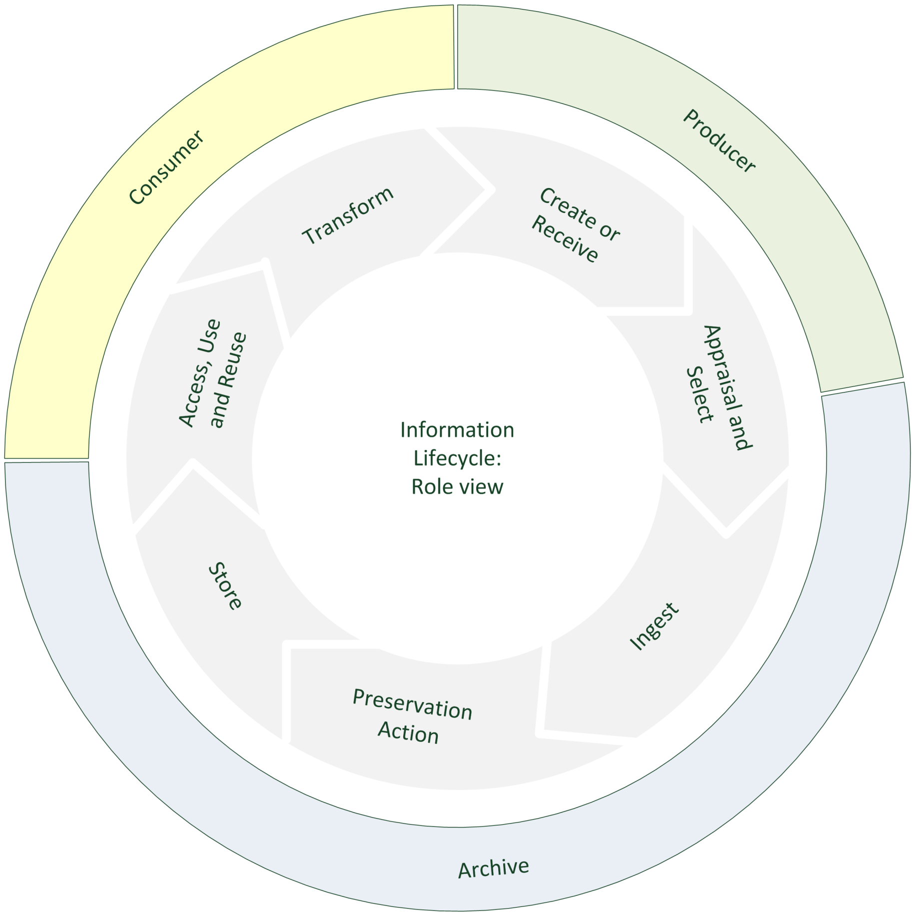


Figure 8: DURAARK Lifecycle

The lifecycle model, shown in Figure 8 for 3D objects in the DURAARK context was already introduced in previous project Deliverables. Hence, it is consequent to reuse this model also for the DURAARK policies.
Beside of that, the entry for 'stakeholders' will stem from the in Deliverable D2.2.1 defined ones:

- Architects and Engineers
- Construction companies
- Researchers and Lawyers
- Building Owners and Real Estate Managers
- Public Administrations, Public Planning and Policy Makers

- Knowledge base maintainers
- Cultural Heritage Institutions

5.3.1 DURAARK Preservation

The Guidance Policy 'DURAARK Preservation' was developed based on the sample policies of the SCAPE Policy Framework and upon information that have been gathered with the help of the questionnaire which was conducted for Deliverables D7.3. Furthermore, this was evaluated again with the help of a questionnaire in Deliverable D7.4. Furthermore they were developed upon the sample policies which have been provided by the SCAPE Policy Framework.

Authenticity	
Related Guidance Policy	DURAARK Preservation
Definition / Description	The APARSEN project stated that from a *"legal point of view, authenticity consists of the capacity of proving the imputability [sic!] of a digital object (generally a record) to the specific person responsible for its creation. In a context of civil law, an object/record is considered authentic when its author (that is its provenance or origin) is undoubtedly recognized."*[14] Since it is not possible to preserve every aspect of a digital object - in fact, in the most cases a file is already an authentic copy of the very original one - the digital archive has to establish its trustworthiness for the user and document any change which might occur through preservation actions or updates. This can be solved with a complete provenance trail and / or appropriate metadata.
Why	For legal reasons it is necessarily required to verify that the preserved material is authentic or at least an authentic copy / migration result of the very original file. However, this is not only important for lawyers, as for instance Architects, Building Owners or Construction Companies need for retrofitting or new planning authentic files as inaccuracies may cause false plannings.
Risks	Without authenticity the trust in the preserved materials may be lost and in consequence also the trust in the digital archive.

Life Cycle Stage	Access, Use and Reuse Ingest Preservation Action Store
Stakeholder	**Architects and Engineers** need authentic files for retrofitting and planning of buildings. Not authentic files disable correct realization of these purposes. **Construction companies** need - similar to 'Architects and Engineers' - authentic files for retrofitting and planning of buildings. Not authentic files disable correct realization of these purposes. **Researchers and Lawyers** want to understand the whole building cycle with all alterations, additions or even destruction. They need therefore authentic files or at least authentic documentation about the lifecycle. Besides, it is important for legal issues to maintain well documented information about the provenance of the digital material. **Building Owners and Real Estate Managers** have interest in sale particular buildings or to plan retrofitting of buildings. **Cultural Heritage Institutions** have the main goal to preserve their material in authentic form and to keep at least information of any changes to reproduce the provenance trail.
Cross Reference	Metadata, Processes and Provenance
Examples	Yale University Library on 'Authenticity': *"YUL strives to ensure the authenticity of digital resources; the mutable nature of digital resources opens the possibility for unauthorized and undetectable changes. Confidence in the authenticity of digital resources over time is particularly crucial owing to the ease with which alterations can be made. From the moment that digital resources are created or acquired, YUL undertakes protective procedures to prevent, discover, and correct loss or corruption of digital resources due to either inadvertent or malicious intent. In addition, supporting evidence, ideally in the form of metadata, must be provided to enable users to evaluate the authenticity of all preserved digital resources."*[21]

Control Policy	All preservation events MUST be recorded. Information on preservation events SHOULD use the PREMIS schema. Information about date, involved persons and action SHOULD be recorded. Checksums MUST be created for the specific file. (algorithm shall be defined as well) Checksums SHOULD be run during ingest. Checksums SHOULD be conducted in intervals. Ingest checksums SHOULD the same as recalculated ones.
Questions to foster discussions	Does your organization record information about the provenance of the object? Do you ensure integrity over the long term with the help of checksums? How do you establish trustworthiness within your organization and especially for the user? Do you track every change that was made?

Metadata	
Related Guidance Policy	DURAARK Preservation
Definition / Description	The simple phrase 'metadata is data about data' get to the heart of a suitable definition. However, metadata is a core issue of digital preservation, as the digital objects provide metadata as well as the digital preservation process. Metadata is needed for the retrieval of the digital material, for its sustainable understanding and required for the entire preservation process.
Why	As stated above, metadata is needed for the preservation itself but also for a meaningful understanding of the preserved material, for instance as representation information. And the most familiar, descriptive metadata is needed for the search and retrieval of the
Risks	Without appropriate metadata preservation is nearly impossible: • The preserved material cannot be retrieved. • The preserved material cannot be understood. • Information about provenance, preservation events etc. cannot be recorded.
Life Cycle Stage	Create and Retrieve Ingest Preservation Action

Stakeholder	**Architects and Engineers** need information about the building they want to retrofit or rebuild, for instance the geo coordinates, address or previous architects. At least a minimal set of descriptive information is needed. **Construction companies** need information about the building they want to retrofit or rebuild, for instance the geo coordinates, address or previous architects. At least a minimal set of descriptive information is needed. **Researchers and Lawyers** need information for the search and retrieval of the architectural data and for the understanding of the building's history. Sufficient descriptive information about the buildings should be provided. **Building Owners and Real Estate Managers** want to retrofit or sale particular buildings. They need detailed descriptive information about the building. **Public Administrations/ Public Planning / Policy Makers** may ask for preserved material as it is required for own retrofitting purposes or knowledge acquisition. At least some descriptive metadata is needed. **Knowledge base maintainers** make architectural data available to various actor and need therefore detailed information about the buildings. **Cultural Heritage Institutions** are prompted to preserve the building data over the long term and have therefore gather descriptive, technical / structural and administrative information about the object. Additionally, every change that might occur have to be recorded within the DPS.
Cross Reference	Authenticity, Processes and Provenance, Standards, Rights Guidance Policy 'SDA / SDO management and interactions'

Examples	Yale University Library on 'Metadata': *"Metadata is fundamental to preserving Yale University Library's digital resources. Preservation metadata includes a number of different types of metadata: administrative (...), technical (...) and structural (...). Particular attention is paid to the documentation of digital provenance (...), and of relationships among different objects within preservation repositories (...)."*[21]
Control Policy	At least every mandatory descriptive element MUST be filled out. At least every mandatory technical element MUST be in place. Every optional descriptive element SHOULD be filled out. Every optional technical element SHOULD be in place. For the file format a Pronom Unique Identifier SHOULD be in place. Information about any event SHOULD be recorded. Every metadata SHOULD be recorded in a suitable standard format. (PREMIS, buildm, ifcm, e57m) A persistent identifier HAS to be in place.
Questions to foster discussions	Which kind of metadata do you capture? Do you store the metadata within the file, along the file and / or in a separate database? Do you use persistent identifiers for sustainable retrieval of the stored objects?Where do you store the metadata? Do you require a minimum set of metadata? Do you consider metadata which is already part of the digital object, like for instance embedded descriptive metadata or technical metadata? Do you use metadata standards?

Preservation Planning	
Related Guidance Policy	DURAARK Preservation
Definition / Description	Preservation Planning was at first mentioned by the OAIS reference model and is one of its central requirements as it ensures the long-term availability of the preserved material. Preservation Planning is build on the services Technology Watch and Community Watch and before a preservation action an extensive and well documented planning process has to be conducted. The success of a preservation action is measured against to be defined requirements.
Why	Since technologies and requirements of the community may change over time, preserved materials cannot be accessible and preserved over time without necessary preservation actions, like for instance emulation or migration.
Risks	Without reacting on changing technologies ans requirements, the accessibility of the preserved material is not guaranteed over time.
Life Cycle Stage	Ingest, Preservation Action
Stakeholder	**Cultural Heritage Institutions** have to ensure the availability, the access and usage of their preserved materials. This is only possible detailed and
Cross Reference	Processes and Provenance Rights
Examples	The National Records of Scotland on 'Preservation Planning': *"Ensure continued access to born digital records by monitoring file formats and their environment and by taking appropriate action. Dertermine most appropriate format for access copies."*[12]

Control Policy	All preservation events MUST be recorded. Information on preservation events SHOULD use the PREMIS schema. Information about date, involved persons and action SHOULD be recorded. Checksums MUST be created for the specific file. (algorithm shall be defined as well) Technical metadata SHOULD be in place. Requirements (significant properties, policies etc.) for preservation planning process MUST be in place. Storage media SHOULD be changed from time to time.
Questions to foster discussions	Are you familiar with disaster recovery strategies, backups, redundant storage at (geographically) different locations, monitoring system for storage media and / or Migration for storage media (exchange of HDD etc.)? Do you have services in place for the identification of technological changes which may affect the access and renderability of your stored data? Do you check, whether a file contains the same content, when you open it with a newer version of the software? Do you store the master file, derivative / working copies of the master files or new created based on the master files or everything which was created? How do you handle files, which cannot be rendered with any common software?

Processes and Provenance	
Related Guidance Policy	DURAARK Preservation
Definition / Description	Within a DPS several processes concerning the preserved material are conducted as for instance the ingest, access or preservation planning. Since digital preservation is a late stage in an object's lifecycle, several changes, editions and additions are probably made before the ingest into the digital archive which is why this provenance information should be documented as well. In particular for architectural data it is important to tell a story and the context from the building for a broader understanding.
Why	The documentation of all done editions and changes is necessary for the authenticity of the trustworthiness of the preserved materials. Even if a digital object was altered several times it can be seen as authentic if all changes are decent documented.
Risks	Without documenting all processes and alterations the authenticity may be lost which leads to a loss of trustworthiness of the digital archive. Additionally a sustainable understanding of the original context is not possible
Life Cycle Stage	Appraisal and Select Preservation Action
Stakeholder	**Architects and Engineers** need detailed information about the provenance of architectural data as well as about every process which was conducted on the data. **Construction companies** need detailed about the provenance of architectural data as well as about every process which was conducted on the data. **Researchers and Lawyers** have a high interest in detailed information about the provenance of architectural data. Besides, it is important for legal issues to maintain well documented information about the provenance of the digital material. **Cultural Heritage Institutions** have to gather and preserve any information about the provenance of the architectural data as well as every process which was conducted.

Cross Reference	Authenticity, Preservation Planning, Metadata
Examples	This Preservation Procedure Policy is highly specific and can be seen as addition to the presented policies on 'Authenticity, 'Preservation Planning' and 'Metadata'. However, for DURAARK purposes it was of high importance to have a policy for the Preservation Procedure in place.
Control Policy	All preservation events MUST be recorded. Information on preservation events SHOULD use the PREMIS schema. Information about date, involved persons and action SHOULD be recorded. All created information during ingest SHOULD be recorded. All created information during access SHOULD be recorded.
Questions to foster discussions	Does your organization record information about the provenance of the object? Do you track every change that was made? Do you collect metadata on the creator, the creation context and the time? Do you have a log file in place?

Standards	
Related Guidance Policy	DURAARK Preservation
Definition / Description	When it comes to digital preservation, standards plays a big role, as they are supporting the sustainability of a DPS. They can cover storage media, metadata, file formats and processes. Even the DPS itself can be build upon a standard. Often standards are certified by organization like ISO[29] or DIN[30].
Why	Standards are often build upon experiences, tests and evaluations and therefore consists on knowledge which can be reused by an institution without providing resources for own developments. Additionally, if standards are certified and internationally accepted they foster the exchange with other institutions and it is possible to benefit from their experiences as well, when it comes to changes of the standards.
Risks	Without using standards, an institutions has to put a lot more resources and efforts in its own digital preservation activities. Additionally, an institution may easily be in an isolated position since there are no possibilities for exchange.
Life Cycle Stage	Create and Retrieve Preservation Action Access, Use and Reuse

[29] http://www.iso.org/iso/home.html
[30] http://www.din.de/de/

Stakeholder	**Architects and Engineers** want to plan new or retrofit existing buildings and are partly depended existing architectural data. To allow a trouble-free subsequent processing, a open and free standard is recommended. **Construction companies** want to plan new or retrofit existing buildings and are partly depended existing architectural data. To allow a trouble-free subsequent processing, a open and free standard is recommended. **Knowledge base maintainers** want to reuse and spread information about stored architectural data for multiple stakeholders. A independent understandable metadata format supports a better accessibility of these information. **Cultural Heritage Institutions** are always invited to use standards or at least recommended formats for the storage of data and metadata. This applies for architectural data as well.
Cross Reference	Metadata
Examples	OAIS reference model, METS, PREMIS, Dublin Core GESIS on 'Metadata': *"The Data Archive embraces the use of recognized standards in fulfilling its tasks (...). Particularly important to planning, implementing, and developing digital preservation is the Open Archival Information System (OAIS) reference model. (...) Metadata, critical to the long-term preservation, documentation, and use of digital data, are assigned according to recognized standards (...)."*[15]

Control Policy	Every metadata SHOULD be recorded in a standardized format. Information on preservation events SHOULD use the PREMIS schema. All digital material SHOULD be preserved in standardized file formats. The used storage media SHOULD be standardized. Probably used interfaces SHOULD be standardized. If no standards are in place, an institution SHOULD orient on best practice solutions.
Questions to foster discussions	Which kind of metadata do you capture? Do you store the metadata within the file, along a file and / in a separate database? Do you require a minimum set of metadata? Do you use persistent identifiers for sustainable retrieval?

Access	
Related Guidance Policy	DURAARK Preservation
Definition / Description	Access provides direct usage of the preserved material. This can be solved with the help of web interfaces or client side software. Access can also provide the transformation of preserved materials in access copies with reduced metadata or low resolution files. If needed it provides additional information about the structure and file format for rendering the provided files.
Why	Access is needed if an digital archive wants to present its preserved material for the public.
Risks	Without provided access to the preserved material the data cannot be used through external stakeholders.
Life Cycle Stage	Access, Use and Reuse

Stakeholder	**Architects and Engineers** urgently need access to the preserved material since it is a necessary starting point for retrofitting activities. **Construction companies** urgently need access to the preserved material since it is a necessary starting point for retrofitting activities. **Researchers and Lawyers** need access for their research activities. **Building Owners and Real Estate Managers** uses architectural data for planning or retrofit a building. Therefore access to preserved material is needed. **Public Administrations/ Public Planning / Policy Makers** are considered as consumers of architectural data and need therefore access to preserved materials. **Knowledge base maintainers** need access to the preserved material for further processing and reuse of the gathered information. **Cultural Heritage Institutions** should provide the access to their preserved materials. This includes also necessary actions to provide a smooth usage of the material, e. g. preservation action, generation of access copies or at least its delivery.
Cross Reference	Authenticity, Rights
Examples	Yale University Library on 'Access': *"In preserving the accessibility of digital resources, the Library will:* • *Maintain information regarding rights and permissions governing access.* • *Maintain the means of accessing an acceptable presentation of the digital resource; and* • *Maintain the ability to locate the digital resource reliably."[21]*

Control Policy	The user MUST access a copy of the preserved material. For the file format a Pronom Unique Identifier SHOULD be in place. The presented material SHOULD be provided with sufficient information about the creation context and provenance.
Questions to foster discussions	How do you offer your stored materials? (Web interface, client side software etc.) Do you present your material in a specialised design, for instance with reduced metadata or 'low resolution' access copies? Do you offer or provide information with which the user may understand the preserved material

Rights	
Related Guidance Policy	DURAARK Preservation
Definition / Description	Rights can be associated with nearly everything within an institution as they formalize legal constraints when it comes to access, usage or manipulation of data or metadata.
Why	With the help of rights it is clearly regulated, what is allowed, what it forbidden or where are constraints.
Risks	Without rights management required information is missing on what is allowed with regards to the data and what is not allowed.
Life Cycle Stage	Create or Receive Ingest Preservation Action Store Access, Use and Reuse

Stakeholder	**Architects and Engineers** require access for the usage of architectural 3D data, e. g. for retrofitting purposes. Rights maintain, if the access or if the manipulation (for retrofitting) is allowed. **Construction companies** require access for the usage of architectural 3D data, e. g. for retrofitting purposes. Rights maintain, if the access or if the manipulation (for retrofitting) is allowed. **Researchers and Lawyers** require rights for access and usage of the preserved material. Especially for lawyers rights are of high importance as they feature legal constraints and burden of proof. **Building Owners and Real Estate Managers** establish rights management to regulate access, usage and manipulation of their own data. **Knowledge base maintainers** are dependent on rights to know, for which kind of data / knowledge the reusage is permitted. **Cultural Heritage Institutions** need a sufficient rights management to regulate how for instance material from an external producer is handled, if migrations are allowed or if the material is confidential. Therefore agreements have to be determined. Rights management is highly dependent of the particular material and the goals of an institution.
Cross Reference	Metadata, Access, Preservation Planning
Examples	Copyrights, Digital Rights Management, Illinois Digital Environment for Access to Learning and Scholarship (IDEALS) on 'Rights' (here copyright): *"All items deposited in IDEALS are under copyright unless otherwise stated. In most cases, the copyright owner is noted in the full metadata for the item. If the copyright owner is not stated and if the item is not in the public domain, it should be assumed that the item is under copyright, and that the author(s) or publisher is likely to be the copyright owner. If an item is under copyright, a user can use the item under normal fair use doctrine."*[8]

Control Policy	A rights management SHOULD be in place. A submission agreement SHOULD be formulated between producer and curator of digital material. Only authorized agents ARE ABLE to edit, delete or access the preserved material.
Questions to foster discussions	Are guidelines established to regulate access, edition, deletion etc. on the digital material? Could you imagine to establish formal deposit agreements with external producers which ensure on both sides how the data is handled? Talking about rights management, what kind of rights are the most important or have to be considered the most? (e.g. Copyrights, Access rights, Licenses)

5.3.2 SDA / SDO management and interactions

As stated in the Introduction, the European Commission recommended to specify and report policies for *"the SDA management (SDO-SDAS) and interactions between SDA and OAIS system."* A description on how to preserve the data which stem from the SDA was intended for this Deliverable anyhow, the specification of its interactions and its management with the help of policies is consequent and makes absolutely sense in this section. The preservation of the SDA itself is further described in Section 6.2. The following sections features several Preservation Procedure Policies within the Guidance Policy 'SDA / SDO management and interactions' for the implementation of the suggested recommendation by the EU reviewers. In contrast to the policies presented in Section 5.3.1 the row 'Example' was left out within the SDA policies since they are highly specific and no other policies are known for semantic enrichment respectively the preservation of enrichment components. The whole Guidance Policy is strongly connected with the Preservation Procedure Policy 'Metadata' in Section 5.2

Metadata levels about the physical asset and digital object	
Related Guidance Policy	SDA / SDO management and interactions
Definition / Description	For the Semantic Preservation of the archived digital objects the descriptive metadata plays the most important role. The metadata can be extracted from the primary data, it could be added by the user and / or generated during the semantic enrichment processes. The 'buildm' schema is therefore a vocabulary with fixed mandatory and optional elements about the physical asset and the digital object as well. The crawls on the other hand are triples which are aggregated in the so called 'buildm+'.
Why	The descriptive metadata about the preserved objects has to be preserved as well, but since the metadata is on the one hand fixed and described in a XML schema definition and on the other hand more flexible and dynamic, two different approaches have to be addressed in the semantic preservation.
Risks	Without preserving both descriptive metadata, those about the digital object as well as the physical asset, the whole content related context will be lost. Be it that just these about the digital object will be preserved, the whole real world related context will be missing, just describing a particular set of files without the needed context for understanding.
Life Cycle Stage	Appraisal and Select
Stakeholder	**Researchers and Lawyers** need the information about the physical asset as well as the digital object for the purpose of retrieval. **Construction companies** need the information about the physical asset as well as the digital object for the purpose of retrieval. **Building Owners and Real Estate Managers** preserve and offer architectural 3D data for different purpose. But for efficient reuse of their content metadata about the physical as well as the digital object have to be conducted. **Knowledge base maintainers** provide the information and make them accessible. **Cultural Heritage Institutions** for a cultural and historical understanding of the preserved objects.

Cross Reference	Semantic Enrichment of objects
Control Policy	At least every mandatory element is filled out. Descriptive information SHOULD be recorded in buildm / buildm+
Questions to foster discussions	Does your organization store information about the digital object as well as the physical asset (the building, monument etc.) Would the usage of the buildm schema be beneficial for your organization?

Semantic Enrichment of objects	
Related Guidance Policy	SDA / SDO management and interactions
Definition / Description	All preserved objects can automatically be enriched by the SDA / SDO provided services. With a minor set of included metadata (for instance the geo location or the name of a building) other metadata is automatically added to the metadata record.
Why	Successful semantic enrichment will enable a fast and greater indexing of all preserved objects. Additionally the exchange is supported through the relations of the semantic objects.
Risks	Without semantic enrichment an automated indexing of files is not provided which may lead into loss of knowledge and lack of ability to retrieve the preserved files by later users.
Life Cycle Stage	Appraisal and Select
Stakeholder	**Knowledge base maintainers** want to build up knowledge and interlink existing vocabularies for providing access. **Cultural Heritage Institutions** and equivalent consumers of architectural data are interested in well indexed objects for sustainable preservation and access.
Cross Reference	Metadata levels about the physical asset and digital object, Harvesting and storing crawls within the SDA
Control Policy	Every supplied information was tried to be semantically enriched.

Questions to foster discussions	Is the automatic enrichment and indexing of descriptive information a useful scenario in your organization or do you have to index everything on your own? Could you imagine to extend your own stock of data and information by semantic and crowdsourced data?

Harvesting and storing crawls within the SDA	
Related Guidance Policy	SDA / SDO management and interactions
Definition / Description	For a successful semantic enrichment the crawls of the Linked Open Data are needed and therefore harvested and stored within the SDA. The crawl itself contains the information which enriches the primary data automatically with descriptive information. The crawls are stored within the SDA for a sustainable understanding of the enriched metadata.
Why	The enriched buildm+ is stored together with the primary data, and the crawls can be triggered and pushed into the SDAS by a user when required. Exporting crawled data into the triplestore is facilitated to support consequent preservation of selected crawls.
Risks	Without harvesting and storing periodically the crawls, the stored metadata cannot be updated and may be get obsolete.
Life Cycle Stage	Store
Stakeholder	**Knowledge base maintainers** need to harvest and store for the population of an accurate and current knowledge base. **Cultural Heritage Institutions** rely on these harvested crawls since they are a prerequisite for the preservation within an OAIS.
Cross Reference	Semantic Enrichment of objects, Monitoring stored crawls and triggering re-harvests Creating snapshot of SDA crawls for handover to OAIS
Control Policy	Crawls SHOULD be triggered periodically. All crawled data HAS to be stored. All crawled data SHOULD be preserved.
Questions to foster discussions	Do you update your metadata from time to time? Do you crawl metadata and store it in your institution?

Monitoring stored crawls and triggering re-harvests	
Related Guidance Policy	SDA / SDO management and interactions
Definition / Description	Buildm instances can be stored and monitored in the SDAS. A versioning system supports updates to buildm instances. The tool operates on two separate graphs in the SDA, one which is used as a temporary/buffer graph and captures data committed by the workbench, and a final graph where the data is stored 'permanently'. The tool takes the resources from the temporary graph and does a three step process. In the first step it lookups in the final graph whether that resource exists, in case it doesn't it commits the first version of the buildm instance in the final graph, otherwise it adds a new version with an incremented version number. The interlinking between the buildM instances and their revisions is done through the provenance ontology PROV[31], in particular the predicate prov:wasRevisionOf is used to interlink the revision with its buildM instance.
Why	The Linked Open Data contains dynamic concepts and semantic relations which can be altered over time. The alteration means also that the preserved descriptive information might not be valid anymore after some time. Therefore, new crawls have to be triggered for renewed and valid information. Consequently, if buildm instances are updated they are automatically versioned within the SDAS (using the SDA versioning tool).
Risks	Without monitoring the stored crawls, the preserved metadata may might not be valid anymore and the associated architectural data cannot be retrieved.
Life Cycle Stage	Create or Receive Store
Stakeholder	**Knowledge base maintainers** need the monitor function to update their knowledge base.
Cross Reference	Harvesting and storing crawls within the SDA, Handover of new buildm+ records to OAIS-compliant archive

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

Control Policy	Versioning HAS to be conducted All versions HAVE to be stored. All revisions SHOULD be interlinked with the predicate 'prov:wasRevisionOf'.
Questions to foster discussions	Do you align your stored metadata with current knowledge? Do you update youre knowledge base? How are these updates conducted?

Creating snapshot of SDA crawls for handover to OAIS	
Related Guidance Policy	SDA / SDO management and interactions
Definition / Description	Snapshots are being executed by the SDA maintenance and should be prepared as a SIP for the handover to the OAIS. The creation process should be triggered manually or periodically.
Why	The preservation of the whole SDA would be to difficult to realize. However, it can be restored from a RDF snapshot. This snapshot will be therefore transferred into an OAIS compliant archive for preservation.
Risks	Since the SDA is not intended to be a long term archive it may be vanished and so the containing data. All crawled information will be lost which leads to a loss of accessibility of the digital material.
Life Cycle Stage	Appraisal and Select Ingest
Stakeholder	**Knowledge base maintainers** have to create the RDF snapshot to be delivered to the DPS. **Cultural Heritage Institutions** want to preserve the rdf snapshots of the crawled data for a sustainable description of the preserved objects.
Cross Reference	Harvesting and storing crawls within the SDA, Storing SDA crawl snapshots in OAIS-compliant archive
Control Policy	A RDF snapshot HAS to be generated of the SDA. The creation process HAS to be generated either manually or periodically.

Questions to foster discussions	Who creates at which time the snapshot? How is it stored? How is it transferred into the OAIS?

Storing SDA crawl snapshots in OAIS-compliant archive	
Related Guidance Policy	SDA / SDO management and interactions
Definition / Description	The SDA creates a RDF snapshot, will be ingested as one SIP into a DPS. SDA crawl snapshots are preserved within an OAIS compliant archive for sustainable digital preservation and linked to the corresponding objects.
Why	The SDA itself is not an OAIS compliant archive and therefore not sufficient for digital preservation of the crawled and stored material. But since it contains a broad range of metadata - which is not part of the preserved digital objects - at least snapshots of the SDA have to be preserved within an OAIS compliant archive. This can be realized with the help of RDF-snapshots which will be preserved as an own AIP and linked to the corresponding objects. Additionally, the SDA can be restored with the help of those snapshots.
Risks	Since the SDA is not intended to be a long term archive it may be vanished and so the containing data. All crawled information will be lost which leads to a loss of accessibility of the digital material.
Life Cycle Stage	Ingest Store
Stakeholder	**Knowledge base maintainers** have to be informed as data producer about successful preservation of their RDF snapshots. **Cultural Heritage Institutions** ingest and store the received RDF snapshots within an OAIS-compliant archive.
Cross Reference	Creating snapshot of SDA crawls for handover to OAIS Handover of new buildm+ records to OAIS-compliant archive Re-populating an SDA with data from an OAIS-compliant archive

Control Policy	The RDF snapshot HAS to be stored in an OAIS-compliant archive. Knowledge base maintainers SHOULD get a notification on the status of their submitted snapshots.
Questions to foster discussions	Do you store your semantic gathered data within an OAIS-compliant system? How do you preserve your semantic gathered data within the OAIS?

Handover of new buildm+ records to OAIS-compliant archive	
Related Guidance Policy	SDA / SDO management and interactions
Definition / Description	The updated buildm+ metadata will be transferred into the OAIS-compliant archive as an AIP Update of the already stored buildm+ records.
Why	The provenance and relations of the preserved metadata will be captured and the old versions will be stored as well.
Risks	If the the updated buildm+ records are not stored within the OAIS-compliant archive a sustainable digital preservation is only possible for the previous preserved but not for the current ones.
Life Cycle Stage	Preservation Action
Stakeholder	**Knowledge base maintainers** create new snapshots and deliver them into an OAIS-compliant system. **Cultural Heritage Institutions** receive the updated build+ records and ingest and store them within a digital archive.
Cross Reference	Monitoring stored crawls and triggering re-harvests, Storing SDA crawl snapshots in OAIS-compliant archive, Re-populating an SDA with data from an OAIS-compliant archive
Control Policy	New buildm+ records HAVE to be transferred into the OAIS-compliant archive. Existing records HAVE to be updated within the OAIS-compliant archive. Knowledge base maintainers SHOULD get a notification on the status of their submitted snapshots.

Questions to foster discussions	Do you update your already preserved metadata? Which event triggers the renewed handover.

Re-populating an SDA with data from an OAIS-compliant archive	
Related Guidance Policy	SDA / SDO management and interactions
Definition / Description	The OAIS compliant archive provides a DIP created from preserved RDF snapshots. With the help of this snapshots the SDA will be re-populated.
Why	In some cases it might be necessary to build up a new SDA with stored and preserved snapshots from an OAIS compliant archive, e. g. if another institution wants to reuse the generated contents for their own stored building information.
Risks	If this functionality is not enabled the SDA might not re-populate its content after a system failure.
Life Cycle Stage	Create or Receive
Stakeholder	**Knowledge base maintainers** need the preserved RDF snapshot from the OAIS-compliant archive to re-populate the SDA. **Cultural Heritage Institutions** provide preserved RDF snapshots for enable the re-populating of the SDA:
Cross Reference	Storing SDA crawl snapshots in OAIS-compliant archive, Handover of new buildm+ records to OAIS-compliant archive
Control Policy	The queried RDF snapshot HAS to be delivered to the Knowledge base maintainers. If necessary, the OAIS-compliant archive SHOULD provide every preserved RDF snapshot.
Questions to foster discussions	Do you populate your semantic archive from a back-up?

5.4 Institutional preservation policies

In Section 5.2 the very generic policy approach of the SCAPE Policy Framework was presented. Within Section 5.3 policies were presented which are developed for the main purpose of the DURAARK project - the preservation of architectural 3D data - and there-

fore highly specific. However, the reuse of the DURAARK policies is at least intended for the project's heterogeneous stakeholders of the building and architectural area.
But every cultural heritage institution has particular goals and tasks which have to be achieved, e. g. to satisfy information needs on specific questions or to preserve unique materials. Even though the presented DURAARK policies are highly specific, they probably not cover all aspects that are necessary to fulfill the goals of an institution. Hence, an additional policy is recommended to formulate the specific goals and tasks of an institution. Of course the SCAPE Policy Template in Annex A.2 can be used for this.
For DURAARK, institutional preservation policies are not that relevant in the first place as the project offers a more generic approach the preservation of architectural data. But when it comes to preservation planning, those policies have to be considered as they are probably featuring constraints, recommendations or other information which may have an impact on the preservation of data. In Section 4.2 some organizational input factors were presented which have a direct or indirect influence on the selection of significant properties. These influence are often defined within an institutional preservation policy.
DURAARK partner TIB Hannover hosts Rosetta as DPS. For TIB's preservation approach a specific policy, entitled 'TIB Preservation Policy', was conducted. This policy clarifies for instance, which collections have to be preserved, who is considered as Designated Community or how digital preservation is realized within TIB. For instance the DPS of TIB preserve for selected objects a digital object identified or uniform resource name along the file - additional metadata elements which have to be considered and preserved through any preservation action.[16]
However, TIB's policy shows again a subjective view on how digital preservation is influenced by heterogeneous factors.

6 DURAARK Sample Preservation Plan

The previous chapters outlined the processes and requirements which have to be in place when it comes to preservation planning. This section builds upon those definitions and features a sample preservation plan. As stated before, the main focus of this deliverable is to conduct a sample preservation planning for IFC files which is covered in Section 6.1 in an example scenario. The second scenario builds upon the policies which have been described in Section 5.2 and describe a bit preservation of RDF snapshots within a DPS. Since every approach requires different preliminary work that have to be conducted, this section features the two following scenarios:

- **File Level - Migration to new IFC schema:** Preservation Planning is often concerned with the migration or emulation of files, when for instance they are obsolete or not supported by a certain community. Hence, the migration of existing IFC files from a lower into a newer version is a suitable scenario for the most common preservation planning. This approach is described in Section 6.1.
- **SDA preservation:** The SDA itself is not a digital archive but holds and provides data which is necessary for the search, retrieval and understanding of the preserved material. Since DURAARK's focus does not lie on the digital preservation of semantic web resources, an approach of bit preservation of RDF-snapshots and its updates through delta snapshots is described in Section 6.2.

6.1 File Level - Migration to new IFC schema

In Section 3.3.1 of Deliverable D6.6.1 it was outlined that the IFC format version 4 supersedes the 2x* versions. And in the maintenance workflow, which was presented in Deliverable D2.5, this specific scenario was described in step 6 "Evolution of file formats (migration)". However, it is still supported and not an obsolete format. However it is very probable that the older format version will not be supported any more within the community and / or the used technology in future. For example, an institution may require higher model versions in its policies or a widespread digital tool stops supporting the lower version. This scenario might be possible in the next few years which is why it was suitable for the following section.

This likely scenario was selected as a sample preservation plan for this deliverable. The planning process follows the DURAARK Preservation Workflow which is presented in Section 3.3.

6.1.1 DURAARK requirements

Starting point
As specified in the DURAARK Preservation Workflow, a precise requirement for a preservation planning process has to be stated. In this case the IFC format version 2x3 is still in use and supported but it is possible that this support will be stopped in the near future. The most beneficial way of preservation planning is to react in time, i.e. before the file format (versions) become obsolete. Hence, there is a tangible requirement to transform IFC files of version IFC2x3 in version IFC4.

Sample records
As a prerequisite, some records have to be present within the institution before starting a preservation planning process. This serves as both a basis for a later preservation, and for investigating the format and identifying technical characteristics, tools and dependencies. This is described in detail in Section 4.

Set requirements
As described in Section 4 the most important requirements for DURAARK's stakeholder are the access, usage and edition of the digital objects for legal reasons, retrofitting or planning purposes. They should therefore be accessible and preserved in an authentic form. The outlined significant properties for IFC files that have been evaluated and summarized in Section 4.1.7 should be considered to be the major requirements for the preservation planning process.
Additionally, the defined Guidance Policy 'DURAARK Preservation' with its consisting Preservation Procedure Policies - presented in Section 5.3 - have to be considered as well. As already mentioned in the previous sections, the questionnaire which was conducted for D7.4 has shown that the integrity, authenticity and rendering of the preserved material is of high importance for the stakeholders. Hence, at least the Preservation Procedure Policies 'Authenticity', 'Preservation Planning', 'Processes and Provenance' and 'Access' with their respective Control Policies determine the requirements and have to be checked during the preservation planning process.

Thus, the following Control Policies have been taken into account (they are all standard within Rosetta and cane therefore seen as checked):

- Authenticity
  - All preservation events MUST be recorded.
  - Information on preservation events SHOULD use the PREMIS schema.[32]
  - Information about date, involved persons and action SHOULD be recorded.
  - Checksums MUST be created for the specific file. (algorithm shall be defined as well)
  - Checksums SHOULD be run during ingest.
  - Ingest checksums SHOULD the same as recalculated ones.
- Preservation Planning
  - All preservation events MUST be recorded.
  - Information on preservation events SHOULD use the PREMIS schema.
  - Information about date, involved persons and action SHOULD be recorded.
  - Checksums MUST be created for the specific file. (algorithm shall be defined as well)
  - Technical metadata SHOULD be in place.
  - Requirements (significant properties, policies etc.) for preservation planning process MUST be in place. *(Described in Section 4)*
- Processes and Provenance
  - All preservation events MUST be recorded.
  - Information on preservation events SHOULD use the PREMIS schema.
  - Information about date, involved persons and action SHOULD be recorded.
  - All created information during *ingest* SHOULD be recorded.

[32]In fact, Rosetta uses the Ex Libris development DNX which is based on and compatible with PREMIS.

  - All created information during *access* SHOULD be recorded.
- Access
  - The user MUST access a copy of the preserved material.
  - For the file format a Pronom Unique Identifier SHOULD be in place.
  - The presented material SHOULD be provided with sufficient information about the creation context and provenance.

All Control Policies can be seen as checked for a preservation planning process and the main control instance for this preservation planning process are the selected significant properties.
In general, the preservation planning process in the DPS can be started after the requirements have been set. However, they might have to be weighted against each other or possibly re-evaluated after a test run of the preservation planning process.

6.1.2 Evaluation within Rosetta

Select preservation set
At first, the preservation set for the preservation planning process have to be selected from the data inventory of the DPS. Rosetta offers automatic selection criteria, for instance a specific format.

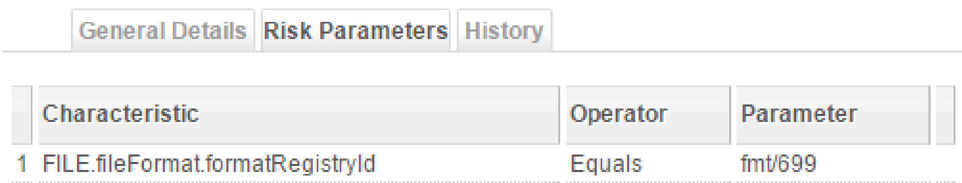


Figure 9: Risk identifier for preservation set selection

This functionality is named as 'Risk Analysis' and enables the user to select different criteria for scanning the total data stock. This is illustrated in Figure 9. In this case the only criterion is that the file format's registry ID - which is based on PRONOM's PUID - equals 'fmt/699' respectively IFC in the format version 2x3.
In theory, more criteria can be added but for the purpose of this test run that seemed

sufficient.

Figure 10: Selected files

After having defined the risk identifiers, the risk analysis can be conducted for the entire data inventory. As shown in Figure 10, all fitting files are presented. Again it is possible to add more conditions for refinement.

Create preservation plan

Based on the retrieved file the preservation plan is going to be created next.

Figure 11: Create new preseravtion plan

At first, some metainformation about the preservation plan has to be provided by the creator. This might e.g. include the name of the plan and the reason to apply it. Additionally, documents can be attached, where for instance the aforementioned identified requirements are described or precise information for the reason of the preservation planning process are provided. In the example shown in Figure 11, the self-descriptive name 'IFC2x3 to IFC4' was chosen. Since it is a test run, this was also provided as a reason. For preservation staff members these information are necessary if they want to recap the reasons and requirements for a preservation plan in the future.

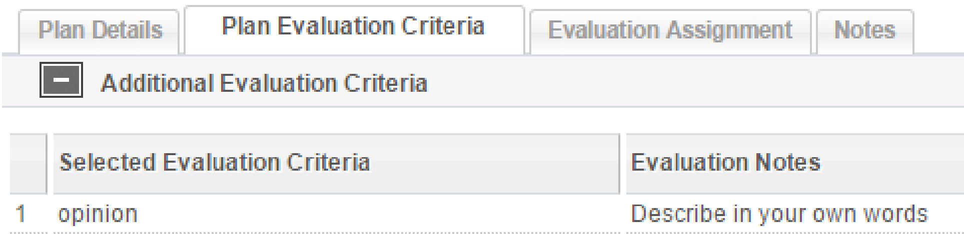


Figure 12: Evaluation criteria for preservation plan

Next is the definition of the evaluation of the basic criteria as shown in Figure 12. The most parts of the assessment is done manually which is why the criterion 'opinion' was selected.

Figure 13: Description of sample preservation action

After the preservation plan was created, one or more plan alternatives can be added. This means, the action which has to be evaluated is added to the plan. In Figure 13 the sample preservation action for DURAARK's planning process is shown. Because there are no suitable converter for a migration from IFC version 2x3 into version 4, the plan will include following test:

1. The identified IFC file of version 2x3 will be exported from Rosetta to a defined server location.
2. The file will be opened in Revit and exported as an IFC file of file version 4. There will be no changes at all.
3. The *ifcm* metadata will be extracted from the newly created file.
4. The new IFC file will be moved into the defined server location and the import will be conducted manually from inside the Rosetta System.

As mentioned in Section 3.3.2, the metadata extractor for IFC files (and for e57 files as well) are not in place in the Rosetta System. This is the reason why the technical metadata will be re-imported as SourceMD.

Test preservation plan
In the next step the whole preservation plan will be performed as externally as described above. The file was exported, converted with the help of Revit and finally imported again into Rosetta.

Evaluate preservation plan
Finally, the preservation plan can be evaluated in the Rosetta system. This can be feasible, if a total automated assessment of the preservation action is possible.

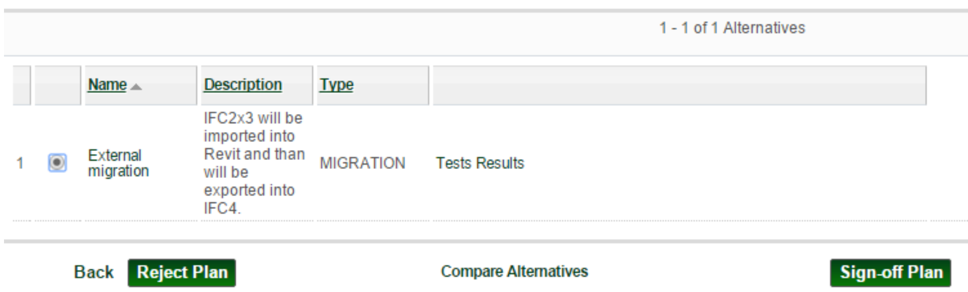


Figure 14: Rosetta internal evaluation of the preservation plan

Figure 14 shows the Rosetta view of the final evaluation. 'Sign-off' would mean that the test run was successful and the plan can be re-used in a productive environment. If the results were not sufficient the plan can be rejected.

However, in case of the manual assessment a detailed evaluation of the plan is more feasible which is why for this sample scenario we let it be as it is.

6.1.3 Recommendation for preservation action

Consider results
At first all gathered results have to be considered and evaluated. This should be conducted with care and in detail. Even though a lot of efforts and resources have been put into the preservation action, the preservation staff should not hesitate to rejecting the whole tested approach as it ensures a sustainable preservation. Additionally, valuable knowledge has been obtained in any case.

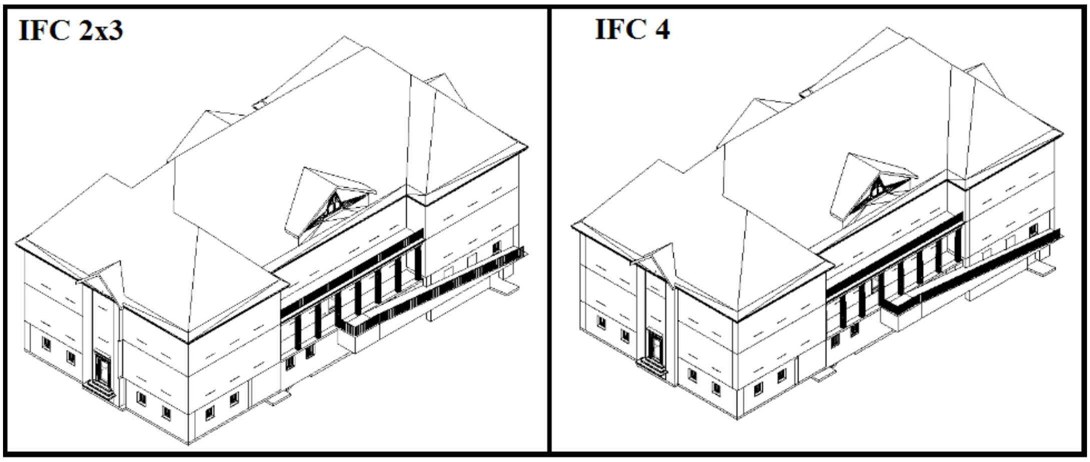
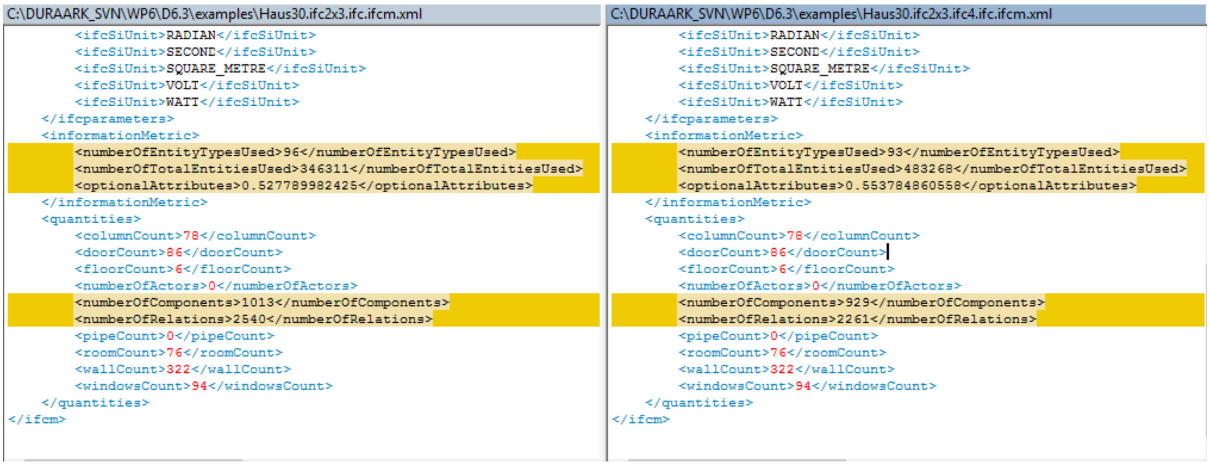


Figure 15: Comparison of visual appearance

The IFC2x3 file as well as the IFC4 file were both executed within Revit and compared against each other. Figure 15 shows on the left side a screenshot of IFC2x3 inside of Revit and on the right side a screenshot of IFC4 inside of Revit as well. Obviously, the windows of the building look a bit strange. However, this is the case in both files and has therefore nothing to do with the preservation action. The grid or fence in front of the building and on the balcony looks on the new created IFC4 file slightly fuller than before.


C:\DURAARK_SVN\WP6\D6.3\examples\Haus30.ifc2x3.ifc.ifcm.xml

```
        <ifcSiUnit>RADIAN</ifcSiUnit>
        <ifcSiUnit>SECOND</ifcSiUnit>
        <ifcSiUnit>SQUARE_METRE</ifcSiUnit>
        <ifcSiUnit>VOLT</ifcSiUnit>
        <ifcSiUnit>WATT</ifcSiUnit>
    </ifcparameters>
    <informationMetric>
        <numberOfEntityTypesUsed>96</numberOfEntityTypesUsed>
        <numberOfTotalEntitiesUsed>346311</numberOfTotalEntitiesUsed>
        <optionalAttributes>0.527789982425</optionalAttributes>
    </informationMetric>
    <quantities>
        <columnCount>78</columnCount>
        <doorCount>86</doorCount>
        <floorCount>6</floorCount>
        <numberOfActors>0</numberOfActors>
        <numberOfComponents>1013</numberOfComponents>
        <numberOfRelations>2540</numberOfRelations>
        <pipeCount>0</pipeCount>
        <roomCount>76</roomCount>
        <wallCount>322</wallCount>
        <windowsCount>94</windowsCount>
    </quantities>
</ifcm>
```

C:\DURAARK_SVN\WP6\D6.3\examples\Haus30.ifc2x3.ifc4.ifc.ifcm.xml

```
        <ifcSiUnit>RADIAN</ifcSiUnit>
        <ifcSiUnit>SECOND</ifcSiUnit>
        <ifcSiUnit>SQUARE_METRE</ifcSiUnit>
        <ifcSiUnit>VOLT</ifcSiUnit>
        <ifcSiUnit>WATT</ifcSiUnit>
    </ifcparameters>
    <informationMetric>
        <numberOfEntityTypesUsed>93</numberOfEntityTypesUsed>
        <numberOfTotalEntitiesUsed>483268</numberOfTotalEntitiesUsed>
        <optionalAttributes>0.553784860558</optionalAttributes>
    </informationMetric>
    <quantities>
        <columnCount>78</columnCount>
        <doorCount>86</doorCount>
        <floorCount>6</floorCount>
        <numberOfActors>0</numberOfActors>
        <numberOfComponents>929</numberOfComponents>
        <numberOfRelations>2261</numberOfRelations>
        <pipeCount>0</pipeCount>
        <roomCount>76</roomCount>
        <wallCount>322</wallCount>
        <windowsCount>94</windowsCount>
    </quantities>
</ifcm>
```

Figure 16: Comparison of the corresponding *ifcm* metadata

While comparing the technical metadata before and after the preservation action it is conspicuous that some properties have changed. In Figure 16 it can be seen that the following properties have changed:

- **ifcm:InformationMetric:numberOfEntityTypesUsed:** was *96* before and is *93* after preservation action
- **ifcm:InformationMetric:numberOfTotalEntitiesUsed:** was *346311* before and is *483268* after preservation action
- **ifcm:InformationMetric:optionalAttributes:** was *0.527789982425* before and is *0.553784860558* after preservation action
- **ifcm:countObjects:numberOfComponents**[33]**:** was *1013* before and is *929* after preservation action
- **ifcm:countObjects:numberOfRelations:** was *2540* before and is *2261* after preservation action

Hence, the deviations are quite explicit even though it seems to have no impact on the visual appearance. However, all listed properties have been considered as significant properties which have to be preserved during the preservation action. At least the preservation action did not change every property.

[33] As mentioned in Section 4.1.2 this was renamend into 'quantities' which is why there is a contrast between figure and text.

Furthermore, an investigation - which was done in the context of Deliverable D7.4 - has shown that the property 'totalWindowArea' was also altered in a very noticeable way: the original IFC2x3 file had a 'totalWindowArea' of '241.211124', the migrated IFC file has '125.1885723'. Even though this property is not part of the *ifcm* schema yet[34], it is an additional tracer that the authenticity of the file cannot be preserved in this preservation action.
Apart from that, there is no explanation as to why the grid or fence looks a bit different. Perhaps, this was just a matter of view adjustments in Revit.

Conclusion
With regards to visual appearance there do not seem to be a deviation between the original and the migrated file. The grit / fence looked a bit different after migration but this might be the result of different adjustments inside of Revit while taking the screenshots. More obvious and crucial are the deviation of parts of the extracted technical metadata. However, it was detected that the parts 'NominalHeight' and 'NominalWidth' of the entity 'IFCWINDOW'[35] inside of the IFC files were changed after the migration. From these parts the property 'totalWindowArea' is computed. Beside of that, the investigations during the process did not lead to a conclusion why the observed deviations occur. The reason may lay in the different schema definitions of the IFC file format, but since this is not sure for a hundred percent (as well as the impact on the authenticity of the file) the preservation plan cannot be seen as successful.
Hence, the following things can be stated as a result of the sample preservation plan:

1. A migration with the help of Revit is not recommended for IFC files since the changes occur which cannot be explained.
2. Further investigation on these changes is needed.
3. A tool for migration of IFC files has to be developed.

[34] A new version of *ifcm* will feature this property.
[35] `http://www.buildingsmart-tech.org/ifc/IFC2x3/TC1/html/ifcsharedbldgelements/lexical/ifcwindow.htm`

6.2 SDA Preservation

To ensure a sustainable preservation, the stakeholders needs not only to know how to keep the files and to render them but also how to understand and interpret the data.

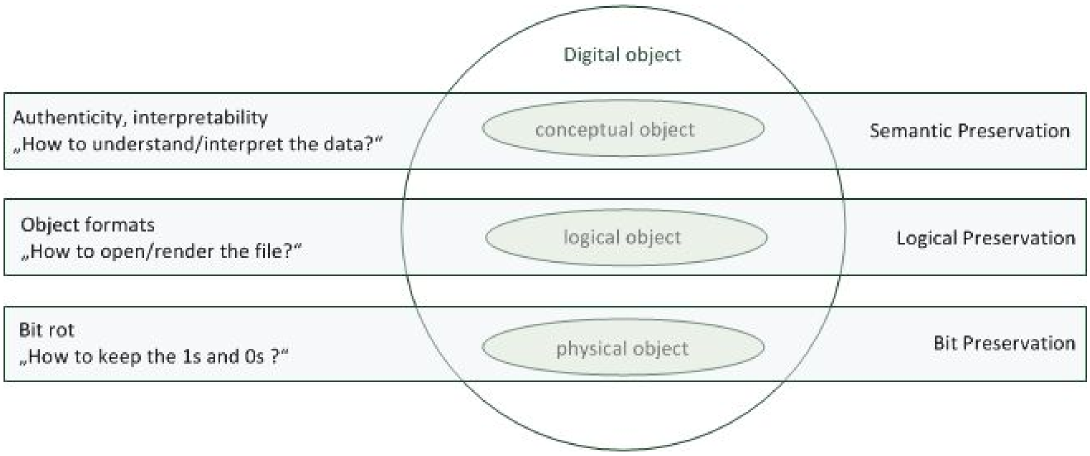


Figure 17: The three layers of a digital object

Figure 17 was first introduced in Deliverable D6.6.1 and shows the three different layers and preservation approaches which have to be conducted for a sustainable digital preservation. While bit preservation is mostly solved through good IT practice nowadays, Logical preservation is realized by different approaches, for instance file identification, validation and also preservation planning. Hence, the processes and developments have been presented in Deliverable D6.2 and in the previous sections of this deliverable are chosen due to their domain specific character. However, even if the bit and logical preservation of architectural 3D data is ensured over the long term, it might not be understandable for its users. Therefore, additional information is needed to allow an independent content related understanding of the preserved material - the semantic preservation.

6.2.1 Semantic Preservation

In the DURAARK project the semantic preservation is partly realized with the preservation of `buildm` records along the file. However, the `buildm` records are only a snapshot of semantic knowledge to a given time. The construction area underlies a constant change, since there are retrofitting activities, new constructions, demolitions or even renaming of

streets or cities. Hence, the stored buildm records are perhaps obsolete after a while and the semantic preservation is not ensured as well as an independent understandability.
Within the DURAARK project, the SDA was established to ensure a dynamic and extensive source for interlinked and current resources. It enables the creation of `buildm` records but has a broader range, which was called 'buildm+' within the project. The SDA offers an approach to ensure the understandability of the preserved objects over time as it builds up a knowledge base that links to the data which is preserved inside the Rosetta system. However, the SDA itself is not an archive in the traditional LTP meaning of the word. It stores the data, but more to enable the semantic enrichment as activity within the project and not to preserve the knowledge over the long term.
With Rosetta, a long term archive is available. The DURAARK WorkbenchUI has already developed the interface between the workbench and Rosetta which enables the ingestion of data. This is documented in Deliverable D2.5 and evaluated in Deliverable D7.4. This module can be reused for the ingestion and preservation of RDF snapshots of the SDA. Currently, only their bit preservation is implemented. This is conducted within a DPS which is why the physical availability is secured over the long term. From time to time - for instance after big changes in the SDA or periodically - new snapshots are created and will be transferred into Rosetta as well as an update to the already existing snapshots.
The preserved snapshots can be used to re-populate an SDA or to build up a new one if the technical infrastructure is in place. Provided the re-population was successful, the preserved data and with it the SDA linked architectural 3D data is extended by a component that enables contextual and content-related understandability. The semantic preservation of the architectural information is thus realized by the bit preservation of SDA snapshots.

6.2.2 SDA bit preservation procedure

In the following a proof of concept, based on the considerations described above is provided. This is strongly connected to the Guidance Policy 'SDA / SDO management and interactions' which was presented in Section 5.3.2.

Prerequisites
First, it is assumed that at least the mandatory elements of the *buildm* record are filled

out before the ingestion process starts in order to have at least some starting points for the semantic enrichment and linking to the SDA. These elements should feature information about the physical asset as well as the digital object. This fulfills the requirements which have been defined in the Preservation Procedure Policy 'Metadata levels about the physical asset and digital object'.

Second, the semantic enrichment of the architectural data was conducted as required in the Preservation Procedure Policy 'Semantic Enrichment of objects'. For starting the enrichment some descriptive metadata is required, for instance the above mentioned mandatory elements. Based on this, the SDA enriches the architectural data from its collected crawls. The enrichment provides opportunities with regards to the Semantic Preservation of the data and additionally serves as an approach of indexing.

Third, it is a precondition that crawls of the Linked Open Data are harvested and stored within the SDA to enable the enrichment of architectural data. This approach serves the requirements which are defined in the Preservation Procedure Policy 'Harvesting and storing crawls within the SDA'.

Fourth, the stored crawls have to be monitored, as there may be additions, manipulations and deletions which might affect the knowledge base with regards to the architectural data. The purpose of the SDA is to serve current and not obsolete data for a sustainable semantic enrichment. Since the Linked Open Data is an dynamic structure a lot of changes may occur. Hence, re-harvests have to be conducted to update the current state of the SDA. This was required in the Preservation Procedure Policy 'Monitoring stored crawls and triggering re-harvests'. In this particular policy the DURAARK approach with the versioning tool is also described.

Creating RDF snapshot

Within the Preservation Procedure Policy 'Creating snapshot of SDA crawls for handover to OAIS' it was defined that snapshots of the whole SDA have to be conducted as RDF file. It is not specified if this is triggered manually or periodically. However, the whole SDA will be exported as N-Quads[36] encoded RDF dataset.

The snapshot looks for instance as shown in listing 1

```
<http://www.w3.org/2002/07/owl#complementOf>
    <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
```

[36] https://www.w3.org/TR/n-quads/

```
        <http://www.w3.org/1999/02/22-rdf-syntax-ns#Property>
        <http://www.w3.org/2002/07/owl#> .
    <http://www.w3.org/2002/07/owl#unionOf>
        <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
        <http://www.w3.org/1999/02/22-rdf-syntax-ns#Property>
        <http://www.w3.org/2002/07/owl#> .
    <http://www.w3.org/2002/07/owl#imports>
        <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
        <http://www.w3.org/1999/02/22-rdf-syntax-ns#Property>
        <http://www.w3.org/2002/07/owl#> .
```

Listing 1: N-Quads encoded RDF snapshot of the SDA

This snapshot should be packaged as a compressed archive such as .tar, .rar etc. in order to be ready for the delivery to Rosetta. Since only a bit preservation of the snapshot will be conducted, no further metadata is required for the delivery. However, it is recommended to at least provide the following information about the snapshot:

- Creation Date
- File name
- Creator / Creating system

Within this proof of concept the snapshot and packaging was conducted manually. It is intended to implement this function within the WorkbenchUI.

Storing SDA crawl snapshots in OAIS-compliant archive

The packaged snapshot will be delivered manually to TIB's environment for ingestion into the Rosetta system. Inside Rosetta, a specific workflow was implemented that connects the ingested data with 'DURAARK' as a producer and which requires some descriptive metadata, like title and description of the package. Since the ingestions from the WorkbenchUI into Rosetta with its created DURAARK workflows worked successfully, an ingest with the new created workflows should work as well.

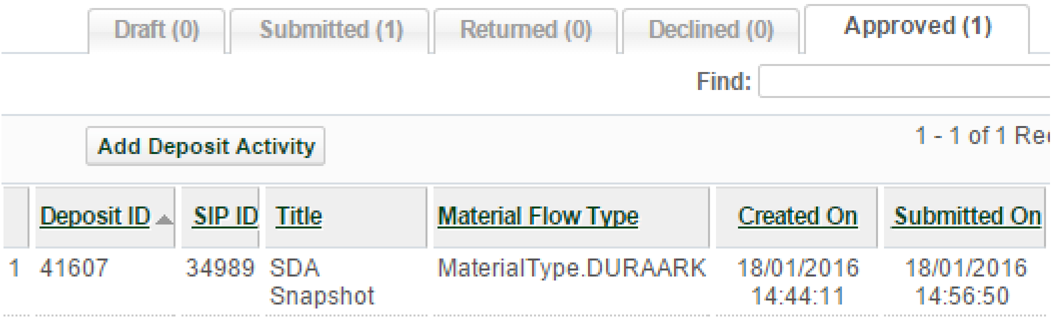


Figure 18: Deposit area of Rosetta with the SDA snapshot marked as 'Approved'

As shown in Figure 18 the data was successfully submitted into Rosetta and approved by TIB's preservation staff. This applies with the restriction, that the file validation failed. This is an expected behaviour, due to the N-Quads encoding which is not known to the PRONOM registry[37] of the National Archives. An identification and continuing validation was not possible. Since this approach features only the bit preservation of the snapshot in the first place, it is unfortunate but does not affect this approach at all.

Handover of new buildm+ records to OAIS-compliant archive

To ensure that the SDA is preserved in the most current state, from time to time new snapshots have to be conducted and delivered into the Rosetta system. The following issues need further discussion and are not finally clarified within this Deliverable:

- How is the event triggered, manually, periodically or after a change occurs?
- Is a delta snapshot sufficient, like it has been discussed in Deliverable D6.2, or has a full snapshot to be conducted?
- Should the snapshot be added as a new representation, like 'Derivative Copy', or ingested as a new package.

Since the whole workflow was conducted manually the handover of the new records were close to the ingest process. The package was delivered to TIB's preservation staff and an update of the existing package was conducted.

[37] `http://apps.nationalarchives.gov.uk/PRONOM/Default.aspx`

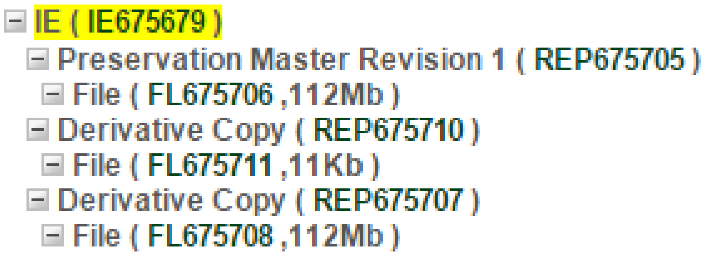


Figure 19: Rosetta view on the SDA snapshot with two updates, each as Derivative Copy

As shown in Figure 19 the original ingest automatically created the Intellectual Entity `IE675679` inside the Rosetta system with the Preservation Master `REP676705` and its containing File `FL675706`. The new SDA snapshot was added as a new representation 'Derivative Copy' `REP675710` with the File `FL675711`. Another Derivative Copy `REP675707` with the File `FL675708`. After this approach any new SDA snapshot will be added as Derivative Copy to the existing record. A new snapshot is not exactly a Derivative Copy, but since Rosetta enables sufficient bit preservation even for Derivative Copies, this approach seems to fit the needs of a proof of concept. However, other approaches should be discussed, for instance in order to ingest a new package every time instead of updating an existing one.

Re-populating an SDA with data from an OAIS-compliant archive
If the SDA does not exist any more or an SDA has to be installed independently from the original one - for instance in a separate infrastructure - the stored SDA snapshots can be downloaded. From the RDF snapshots the SDA can be restored into any NQuad compatible triple store which are commonly available. Given the fact that there are perhaps multiple versions of the SDA, the knowledge and the linked 3D data can be rebuilt dynamically.

6.2.3 Conclusion

As discussed above, preservation of the DURAARK knowledge base ensures the long term availability of the populated SDA and the understandability of the enriched 3D data. However, this is strongly dependent of the user's knowledge regarding how to reuse the snapshots since only bit preservation and not logical or semantic preservation is in place for this proof of concept. Furthermore, some issues remain to be addressed as e.g.

the trigger for new snapshots.
Altogether the described workflow of the SDA snapshot preservation offers a first approach for the long term availability of developed SDA knowledge base.

7 Conclusion and DURAARK Objectives

With the sample selection and definition of significant properties and the sample preservation plan, this deliverable has clearly addressed the open gaps which were identified in D6.6.1. A process for preservation planning was developed based on the best practice work which was conducted in the PLANETS and SCAPE projects and taking Rosetta into account within the DURAARK environment. The defined significant properties for IFC files were selected according to the best practice work of the InSPECT project, taking not only the technical nature of the file into account but also organizational factors that might affect its selection. The presented policies for preservation and SDA / SDO management and interactions were built upon the work which was by the SCAPE project considering the needs and requirements of the DURAARK workflow.
The provided sample preservation plan for IFC files has shown that the chosen preservation action with Autodesk Revit was not successful and that alternative tools have to be evaluated. Furthermore it has to be investigated what has caused the identified deviations. However, even though the action itself was not successful, it verifies that the defined significant characteristics were well chosen constraints for the preservation action.

7.1 Integration Scenarios

The sample preservation plan and the work with regards to planning processes done previously, significant properties and policies have a prototypical character and are therefore not transferable without modification into the productive environment of any stakeholder. However, the provided work can be used as a base for the own selection of significant properties, the formalization of workflows with the help of policies or to conduct a planning scenario with similar focus.
TIB Hannover uses the Rosetta system as the DPS solution in their production environment. Also, the TIB already has several 3D data sets in their archive , which is why the DURAARK foreground results including the WorkbenchUI will be reused. Currently, an implementation of the metadata extractors would fail due to incompatibility issues, but this is solved within the next months. For the assessment during preservation planning and its evaluation the implementation of the metadata extractors would be a great benefit.

7.2 Impact and Outlook

Task 6.4 required a sample preservation plan for 3D objects based upon a clear concept for decision making, defined technical and organizational requirements respectively constraints and an exemplary test run. This task was therefore the major goal of this deliverable and was addressed as summarized above. The in-depth discussion on preservation planning of architectural 3D data with all is facets has shown how to establish an approach for a format that was never addressed in an archival context before. With the sample preservation plan and its negative evaluated preservation action also showed how effective the elaborated approach with its constraints helped during the assessment of the preservation action.
However, this Deliverable mostly focuses on the preservation planning for IFC Files. Even though the preservation planning process and policies presented here are generic approaches (at least for the architectural area), the selection of significant properties and the conducted sample preservation plan are IFC-specific and not applicable for E57 files. Further investigations on possible significant properties and a sample preservation plan for this second file format in the focus of the DURAARK project is therefore strongly recommended.
In Section 3 two other options for preservation actions beside migration were shown. 'Technical museums' were considered as infeasible, which is why it can be neglected in future developments. 'Emulation' on the other hand is a highly promising alternative to format migrations. Since the migration approach failed within the sample preservation, this alternative has also to be tested and evaluated using a sample scenario.

7.3 Collaborations with other projects

During the three years of the DURAARK project a few collaborations with other projects have been conducted:

- **DEDICATE**[38]
  At the final seminar of 'DEDICATE' (Design's Digital Curation for Architecture), the DURAARK project was presented at the University of Glasgow in October 2013. The Royal Commission on the Ancient and Historical Monuments of Scotland

[38] `http://architecturedigitalcuration.blogspot.de/`

and CyArk Europe had a high interest in the DURAARK project and its results. Therefore the contact for exchange and discussion on further developments was established.

- **RADAR**[39]
  The DFG funded project 'RADAR' (Research Data Repository) works on a interdisciplinary infrastructure for research data with regards to availability, preservation and publication. Exchange between DURAARK and RADAR took place on research data management and the development of information infrastructures.

- **MIT FACADE**[40]
  The project 'MIT FACADE' (Future-Proofing Architectural Computer-Aided Design) was from the beginning of DURAARK one of the most important reference projects with regards to long term preservation of architectural 3D data. The MIT requested information from the DURAARK project for extending their work which was conducted during FACADE.

- **IANUS**[41]
  The DFG funded project 'IANUS' establishes functions and services for supporting digital research information within classical studies. After a presentation of the DURAARK project on the 'Kooperation Langzeitzugriff 2015' in June 2015 in Berlin, IANUS requested further information of the DURAARK project with regards to long term preservation of three dimensional. On this occasion exchange was established.

[39] http://www.radar-projekt.org/display/RE/Home
[40] http://libraries.mit.edu/news/facade-project/457/
[41] http://www.ianus-fdz.de/

References

[1] Antunes et al. Shaman reference architecture. Online, Nov. 2011.

[2] C. Becker, H. Kulovits, M. Guttenbrunner, S. Strodl, A. Rauber, and H. Hofman. Systematic planning for digital preservation: Evaluating potential strategies and building preservation plans. *International Journal on Digital Libraries (IJDL)*, Dec. 2009.

[3] C. Becker, H. Kulovits, A. Rauber, and H. Hofman. Plato: A service oriented decision support system for preservation planning. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL'08)*, 2008.

[4] U. M. Borghoff, P. Rödig, J. Scheffczyk, and L. Schmitz. *Long-Term Preservation of Digital Documents: Principles and Practices*. Springer Berlin Heidelberg, 2006.

[5] CCSDS. Reference Model for an Open Archival Information System (OAIS) - Magenta Book, 2012.

[6] A. Dappert, B. Ballaux, M. Mayr, and S. van Bussel. Pp2-d2: Report on policy and strategy models for libraries, archives and data centres. Online, June 2008.

[7] C. Erickson. Storage for the next 5,000 years. Online, Dec. 2015.

[8] Illinois Digital Environment for Access to Learning and Scholarship (IDEALS). *Access and Use Policy*, July 2014.

[9] InSPECT. Deciding factors: Issues that influence decision-making on significant properties. Online, June 2008.

[10] InSPECT. Final report. Online, Dec. 2009.

[11] InSPECT. Inspect framework report. Online, Oct. 2009.

[12] The National Records of Scotland. *The National Records of Scotland and born digital reords - a strategy for today and tomorrow*, Sept. 2014.

[13] nestor Network of Expertise in Long-Term Storage. Guideline for Preservation Planning. Procedural Model and Implementation Version 2.0. Online, 2014. nestor-materials 15.

[14] A. Project. D24.1 report on authenticity and plan for interoperable authenticity evaluation system. Online, June 2012.

[15] N. Schumann. *Digital Preservation Policy. Principles of digital preservation at the Data Archive for the Social Sciences.* GESIS - Leibniz Institute for the Social Sciences, GESIS - Leibniz Institute for the Social Sciences Data Archive for the Social Sciences Unter Sachsenhausen 6-8 50667 Cologne, 1.4.7 edition, June 2013. Overview of principles of digital preservation at the GESIS Data Archive for the Social Sciences.

[16] F. Schwab. *TIB Preservation Policy.* German National Library of Science and Technology, German National Library of Science and Technology (TIB) Welfengarten 1 B 30167 Hannover GERMANY, version 1.0 edition, July 2015.

[17] B. Sierman, C. Jones, S. Bechhofer, and G. Elstrøm. Preservation Policy Levels in SCAPE. In *Proceedings of the 10th International Conference on Preservation of Digital Objects iPRES 2013*, 2013.

[18] B. Sierman, C. Jones, and G. Elstr. Catalogue of preservation policy elements. Electronic, Feb. 2014.

[19] S. Strathmann. *Institutionelle Preservation Policy*, chapter 3.3, pages 6–9. nestor, 2009.

[20] The National Archives. Digital preservation policies: Guidance for archives. Online, 2011.

[21] Yale University Library. *Policy for the Digital Preservation*, Feb. 2007. Original policy established 11/05, updated in 02/07.

Appendix A

Appendix

A.1 SCAPE Guidance Policy Elements

As mentioned in Chapter 5.2 SCAPE provides within the 'Catalogue of Preservation Policy Elements' several Guidance Policies and containing Preservation Procedure Policies which are shortly described in the following.[1]

A.1.1 'Authenticity'

The authenticity of a preserved digital object is crucial since it is according to the OAIS model the evidence for the user that the presented object is at least equivalent to the originally preserved one. Hence, it is necessary for the data owner to prove the authenticity of the provided object.
'Authenticity' itself is not that clear since it depends on the goals and character of an organization but can be achieved through various technical functions and of course organizational policies. This guidance policy with the related procedure policies 'integrity', 'reliability' and 'provenance' offers a generic approach for an institution to secure the authentic preservation of their digital stock.

[1] `http://wiki.opf-labs.org/display/SP/Policy+Elements`

A.1.2 'Bit Preservation'

The basis for a successful digital preservation is to ensure the integrity and readability of all preserved bitstreams which are belonging to one or more digital objects. Functional preservation itself could not be solved without a good bit preservation. Thus, it is important for organization to keep that relationship in mind.

A.1.3 'Functional Preservation'

With the help of functional preservation the permanent access to the preserved material is ensured. This is realized by active intervention if for example technological changes occurs (e.g. new standard formats or new technological requirements). The functional preservation itself requires a careful and well selected respectively tested strategy since it should ensure the access to the digital material for years to come.

A.1.4 'Digital Object'

It is crucial for an institution to define the value of particular collections or digital objects. It has to be clear to preserve the original object over the long term and to keep its accessibility with the help of preservation actions. Therefore, this Guidance Policy is strongly connected with preservation planning, the definition of significant properties and the documentation of all conducted changes or perhaps even deletions of objects.

A.1.5 'Metadata'

Metadata is one of the most important issues when it comes to digital preservation. Without metadata the stored data may just be bitstreams on a server and the broader meaning and understanding of the data will be lost. Hence, without metadata digital preservation is not possible.

There are different kinds of metadata which have to be preserved, because it necessary to outline every needed, mandatory element for a successful preservation.

A.1.6 'Rights'

A good and well documented rights management plays an important role when it comes to digital preservation. With the help of rights the ingest, access, preservation and curation of digital material is managed. Hence, there is a broad variety of rights reaching from the top level - like access into the building where the digital archive is operated - to the very low level - like the rights for editing single metadata field. Beside of that, national law and agreements may have an impact on the rights of the preserved data.

A.1.7 'Standards'

Standards help an institution to re-use experiences and best practice of other institutions and have the state of the art in picture. Thus, is not necessary to develop own standards but link to existing ones. They may be normative or technical and cover mostly any part which can be considered as a component of a digital archive.

A.1.8 'Access'

In most cases the digital preservation has the goal to make the stored data available and usable to the public / the designated community. But since digital preservation is not a matter of just a few years an organization has to consider also possible usage scenarios in the future. Hence, it is necessary to offer the user all information he needs to make the stored data accessible and understandable.

A.1.9 'Organization'

Since digital preservation is not only a isolated part within an organization it has to be supported proactively. This includes every process, the staffing and the budgets which is necessary for a successful digital preservation. The organization itself has to state how to achieve its defined preservation goals.

A.1.10 'Audit and Certification'

For gaining and verify trustworthiness audit and certification is a good and sufficient option. With the help of established options like the Data Seal of Approval, the nestor Siegel in Germany or certification against the ISO 16363.

A.2 Policy element template

DURAARK created its own policies based on SCAPE's policy element template which can be found in the internet[2] and is described in the following.

Preservation Procedure Policy:	**Name of the policy element**
Related Guidance Policy	Because every policy element should be readable independently, the related higher level, the Guidance Policy, is mentioned.
Definition / Description	Every policy element will have a description and, if applicable, a definition, based on existing glossaries in standards like the OAIS model or digital preservation glossaries, such as the APARSEN[3] project or the InterPARES[4] project. The source of the definition will be referenced.
Why	An explanation is given why it is important that an organization defines a policy related to this element
Risks	Not having a written policy could imply various risks for the organization and in this box some examples will be given. Of course whether the risk will occur is dependent on several factors , the examples are added to stimulate further discussions. Apart from general knowledge, also standard literature like DRAMBORA[5] and ISO 16363[6] can be used in these internal discussions.
Life Cycle Stage	The intention of this box is to put the policy element in relation to the life cycle stages it might be relevant for and to achieve a coherence in policy elements for different life cycles. As the basis the DCC life cycle model[7] is used.

[2] http://wiki.opf-labs.org/display/SP/Policy+element+template
[3] http://www.alliancepermanentaccess.org/index.php/knowledge-base/dpglossary/
[4] http://interpares.org/ip2/display_file.cfm?doc=ip2_glossary.pdf\&CFID=4453522\&CFTOKEN=42973887
[5] http://www.repositoryaudit.eu/
[6] http://public.ccsds.org/publications/archive/652x0m1.pdf
[7] http://www.dcc.ac.uk/resources/curation-lifecycle-model

Stakeholder	It is important that someone in the organization will be responsible for describing the preservation policy, in relation to the processes the policy relate to and in coherence with the other processes in the organization. This person is called a "stakeholder". The SHAMAN[8] project distinguished a set of stakeholders in relation to digital preservation and these are used where applicable.
Cross Reference	It is seldom that a policy element stands in isolation. More often a policy element is related to other policy elements, where applicable this relationship is mentioned.
Examples	To illustrate the policy element, one or more relevant examples of Preservation Policies were taken, based on the collected policies. This could be used as an inspiration for organizations to create their own version.
Control Policy	As mentioned before, we have related the control policies to two cases: Preservation Watch and Preservation Planning, as these are the areas in the SCAPE project where the control policies will be used.
Questions to foster discussions	If an organization wants to create preservation policies, it will be important to engage different people in the organization (the "stakeholders") and together phrase the relevant policies. The set of questions for each element will help starting the discussions and highlight the various aspects of the policy element, like the risk of not having thought of the policy element.

[8] http://www.shaman-ip.eu/