A Domain-driven Approach to Digital Curation and Preservation of 3D Architectural Data: Stakeholder Identification and Alignment in the DURAARK Project

Michelle Lindlar; German National Library of Science and Technology (TIB); Hannover, Germany
Martin Tamke; Centre for Information Technology and Architecture (CITA) at the Royal Danish Academy of Fine Arts, Schools of Architecture, Design and Conservation; Copenhagen, Denmark

Abstract

DURAARK is an interdisciplinary EU FP7 project which researches and develops processes and methods for the digital preservation of architectural three dimensional data. Challenges addressed within the project relate to different aspects of the curation and preservation process, including semantic and geometric enrichment, consistent naming schemas and ontologies, as well as pre-ingest tasks in OAIS compliant digital preservation workflows. In a first step, the project identified stakeholders for methods and processes. Since the project strives for a holistic digital preservation approach, different models were taken into account to align the identified stakeholders with the overall curation and preservation workflow. This alignment leads to a better understanding of stakeholder involvement in the process and of the implications this involvement has on preservation decisions. It furthermore outlines the basis for future work in this area.

Introduction

In current digital preservation activities there has been a slow shift from a strong cultural heritage involvement towards a growing awareness in various domains with strong electronic output [1]. The architectural domain of today is undoubtedly dominated by electronic output. Architects benefit greatly from the vast amount of different digital representation forms available, enabling them to overlay three-dimensional descriptions with information on the planning and construction processes. The planning, construction and maintenance of a building involves many players, ranging from architects and engineers during the construction phase to building owners and facility managers during the "operation" phase. Figure 1 shows a typical building lifecycle - the construction phase, which on average lasts about 2.5 years, consists of the stages where the majority of data is produced. After the handover phase only parts of the data produced during the construction phase may be used regularly – here especially the information relevant to parts requiring ongoing maintenance, such as surfaces, lighting or other exposed parts of the building, may be used regularly. Other information, for example information about structural elements or electrial systems, typically only be relevant in the case of re-purposing or emergency. On average, a building is in use for about 60 years before it is demolished or repurposed. Especially in the case of repurposing, the availability of the original data is an important asset. But other scenarios exist, where the long-term availability of building information is of great value, such as in the case of research of historic building structures or city planning.

In the context of preserving the digital output of this domain it is necessary to understand the involved producers as well as potential consumers. Only with an accurate understanding of the stakeholders involved is it possible to meet their needs across curation and preservation activities. The following sections give a brief insight into related work and describe how the current 3D object processing of architectural data of today is a multi-faceted process involving numerous stakeholders. The involved stakeholders are categorized, described and linked to object-centric and process-centric digital preservation and curation models. In a last section, the results are linked to future work to be addressed by the DURAARK project.

Background and Related Work

Despite the high electronic output of the architectural field, few projects have dealt with the digital curation and preservation of domain innate objects and knowledge. The MIT FAÇADE (Future-proofing Architectural Computer-Aided Design) project dealt specifically with the capture, description, management,

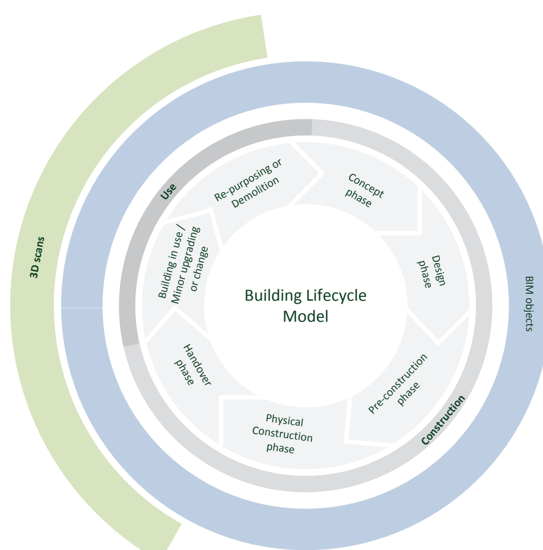

__Figure 1.__ A Building Lifecycle Model: While BIM objects are created, used and updated along the entire lifecycle, 3D scans may document the "as built" state of an erected building.

preservation and availability of architectural CAD models. While the project put a focus on the capture of an entire architecture/design project, a standard format for BIM had not widely established itself during the project running time (2006-2009). Instead, FAÇADE focused on the original CAD file preservation and prototyped software emulation in the digital repository - a technically and legally difficult process [2]. Other domains dealing with 3D data have established ongoing initiatives, the most prominent one being the aerospace and defense industries' LOTAR initiative [3]. In February 2013 DURAARK (Durable Architectural Knowledge) was launched: an interdisciplinary EU FP7 project which addresses gaps in the digital curation and preservation processes and methods of architectural 3D data. The project deals with two kinds of architectural data: "as-planned" data and "as-built" data. While "as-planned" data was formerly available as 2D site plans, the first 3D drafting programs became available in the early 1980s. Over the course of the last few years there has been a paradigm shift within "as-planned" data towards 3D Building Information Models (BIM) – highly annotated models containing detailed information produced and used at the planning, construction and maintenance phase of a building. The pure "as built" state of a building can today be documented through point cloud scans; the tremendous progress made in the availability of various methods over the course of the past 15 years has favored the use of 3D scanning technology in many domains. To meet the problems of legally and technically challenging preservation processes based on legacy formats, DURAARK is researching and developing processes and methods based on two openly available and well-adopted data formats. For BIM objects, DURAARK is focusing on the IFC-SPF file format, which is widely supported by current day CAD software. For point-clouds, the openly available E57 file format was chosen. Both formats fulfil sustainability factors from a digital preservation point of view [4].

3D object processing in architectural practice of today

In current planning and construction practice, the physical object is described through multiple digital objects, which undergo various cycles of active use, conservation and enrichment. Usually organized through a high-level model of the spatial layout (architectural model), each of these digital objects is highly specific to the respective needs of the (sub)domain in which context it was produced – while a model describe a building's façade may disregard the interior, a second model describing the electric layout including panels and switches may disregard the façade. Recently, there has been a global slow but steady shift towards Building Information Modeling (BIM) becoming the predominant approach towards architectural production. BIM objects are a set of 3D-objects which are conceptually able to include the whole range of architectural representations in a single file. These objects are hence introducing a new depth, width and length to the digital representation layer:

- Depth: a BIM is a single digital object which can describe all scales of architecture simultaneously
- Width: a BIM is a single digital object which connects the stakeholders and fields of knowledge in a building simultaneously
- Length: a BIM is a single digital object which links descriptions of all stages of architectural production, of the existing design and of simulations of the future [5]

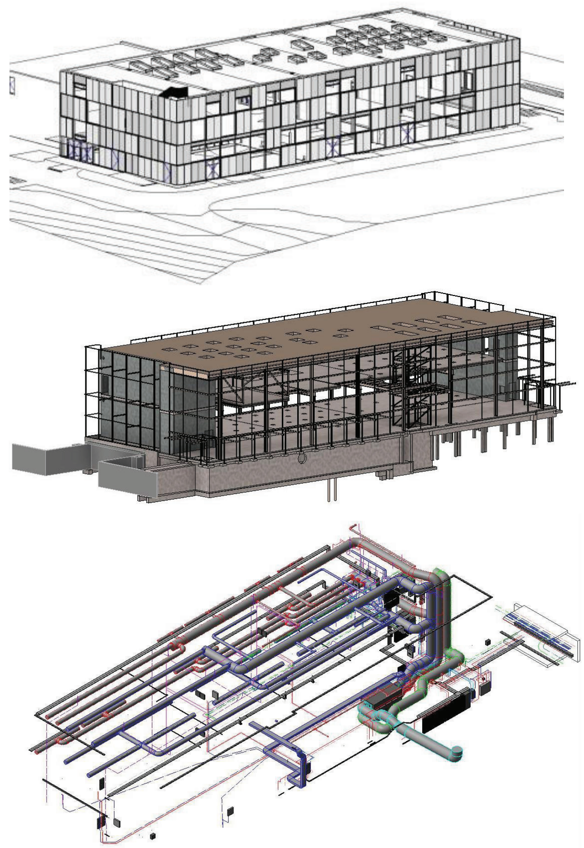

Figure 2. Different views of the same object – architect (top), construction engineer (middle), HVAC engineer (bottom)

While BIM objects allow the detailed documentation of the planned state, pointcloud scans allow a precise documentation of the built state. As such, continuous scans may allow a detailed documentation of the slightest deviation during the "use" phase of a building's lifecycle. Further uses for 3D scanning include spatial analysis, contributing to a record before renovation or contributing to three-dimensional models, animations or illustrations which may be used in visitor centers, museums or on the web [6].

Stakeholder Identification

As mentioned in the previous section, BIM objects introduce a new depth, width and length to the representation layer. These objects include the data of various stakeholders involved in the construction and use/maintenance phase of a building. While 3D scan data only holds the information of one stakeholder, it may be produced for different reasons at various points in a building's lifecycle, therefore supporting different stakeholders' use cases or needs. Questions regarding quality and completeness of the data as well as questions revolving around significant characteristics, i.e., those criteria which shall be kept over the course of preservation action, highly depend on the context in which a digital object was created as well as the scenarios in which the same object shall be used in the future. Given the multi-faceted nature of 3D architectural objects, a domain specific approach cannot consider the entire architectural domain as a singular stakeholder, but needs to take a differentiated look at the various sub-domain stakeholders involved. Hence knowing who these stakeholders are and whether they function as data producers and as consumers – or only as one of the two roles – plays a pivotal role in the context of digital curation and preservation.

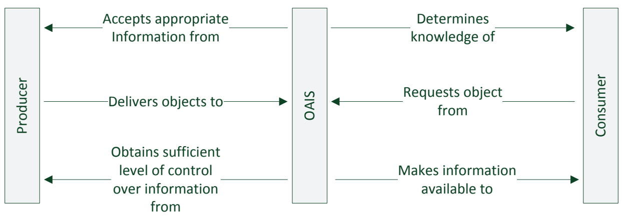

Figure 3. *Basic forms of Producer – OAIS – Consumer interaction in an OAIS-compliant archive*

The OAIS (Open Archival Information System) reference model differentiates between producer, OAIS and consumer, pointing out that the consumer may be – but doesn't have to be – synonymous with the producer. An archive has to fulfil some basic interaction with producers and consumers in order to be compliant with the reference model. As such, it is the archive's responsibility to accept information in an appropriate form and to obtain a sufficient level of control over it. Furthermore, the archive needs to determine the knowledge which can be assumed for the designated community, i.e., the consumer, of the information [7]. Within the DURAARK project, the identification of stakeholders was conducted through an analysis of a building's life-cycle (see Figure 1) and the identification of parties involved in the different stages of the life-cycle. The project identified the following main stakeholder groups as relevant parties in the preservation lifecycle of architectural 3D data [10]:

Architects and Engineers: Architects and engineers are the typical producer of 3D architectural data. They create, update and consult objects during the entire construction cycle of a building. Furthermore, they may act as consumers at the reconstruction stage of a building, where older documentation may be consulted for reference. However, stakeholder interviews with architects and engineers have shown that their interest in long-term archival typically hardly extends the objects initial creation and usage phase for planning purposes.

Construction Companies: During the physical construction phase, construction companies rely on the information provided by architects and engineers. Furthermore, they produce data and/or enrich existing BIM objects with information such as vendor information about building parts. This practice is especially seeing an increase, as the level of detail (LOD) in modern day BIM covers objects as small as 5 mm. Construction companies are increasingly including this form of detailed specification through enrichment in their business models. The contractual situation requires these stakeholders to revisit the buildings regularly for building checks - seemingly on the base of 3D objects. The interest in the long-term archiving of 3D objects covers a digital object's lifespan of 5 years minimum.

Building Owners and Real Estate Managers: After the "handover phase", which marks the end of the construction segment of the cycle, the stewardship over the data often passes from those stakeholders involved in the planning and construction processes to the stakeholders responsible for the maintenance while the building is in use. Even though the "use" segment is the longest phase of a building's lifecycle, very little data is produced during this time. Usage frequency of the data varies based on building complexity and required maintenance for e.g., electrical or HVAC parts. Other usage scenarios include minor reconstruction, risk assessment or value assessment. Owners and real estate managers, as the classical long-term consumers of architectural 3D data, heavily rely on the long-term accessibility and understandability of the data received during the handover phase. As in the case of large facility management companies who hold thousands of records on hand, this stakeholder group may function as an archive as well.

Public Administration / Public Planning / Policy Makers: Public administrations and policy makers can have a strong impact on data production, as a recent shift towards policies which specifically request BIM objects for publically funded buildings can be observed.[1] While these stakeholders influence the way in which data is produced, they typically do not function as a producer themselves. Rather, public planning may have an interest to exploit available data, i.e., for urban planning, which hence positions them at the consumer side of the data lifecycle. Besides the consumer role through exploitation, the public sector may often function as an archive, where the deposit of architectural 3D data – especially in the case of publically funded buildings – may be deposited to a digital archive.

Knowledge Base Maintainers: Currently the knowledge base field is largely dominated by vendors or large organizations. Examples for currently existing knowledge bases are the buildingSMART Data Dictionary (bsDD)[2] or the NBS National BIM Library of the National Building Specification, UK[3]. Such knowledge bases provide generic and/or proprietary BIM objects, such as wall claddings or windows, which may be embedded in larger BIM objects. Knowledge base maintainers typically produce their own data and may function as an archive for their own data.

Cultural Heritage Institutions: Cultural heritage institutions, such as libraries, archives and museums, are often responsible for information that has left the domain of industrial interest and needs to be preserved as part of the cultural heritage of a specific country or region. While cultural heritage institutions are typically archive maintainers, examples like CyArk[4] or RCAHMS – the Royal Commission on the Ancient and Historical Monuments of Scotland[5] – may also produce architectural 3D objects, mainly in the form of 3D scans. As the cultural heritage interest in documentary digital objects of a structure may far surpass the actual lifecycle of the erected structure itself, cultural heritage

[1] See for example the introduction of BIM in Denmark, pushed by the Danish government initiative Det Digitale Byggeri (Digital Construction): `http://www.detdigitalebyggeri.dk/`

[2] `http://www.buildingsmart.org/standards/ifd/dictionary-international-framework-for-dictionaries-ifd`

[3] `http://www.nationalbimlibrary.com/`

[4] `http://archive.cyark.org`

[5] `http://www.rcahms.gov.uk`

institutions should be regarded as a separate consumer.

Researchers and Lawyers: Researchers and Lawyers form a separate stakeholder group, which solemnly functions as a consumer. While lawyers will typically act in connection to another stakeholder group, e.g., building owners, they form a separate designated community with a different knowledge and different requirements in the data's appropriateness and completeness. Research, on the other hand, may act without connection to another stakeholder group and query the data in connection with e.g., socio-economical or historical research questions, therefore introducing a new set of knowledge connected to – or already contained within – the architectural 3D object.

Stakeholder Alignment with Curation and Preservation Processes

In digital curation and preservation discourse, two models are frequently used to align processes and methods. The first model is that of the multi-layered object with its corresponding preservation layers, based on observations made by Thibodeau [8]. The layer model can very well function in a stakeholder agnostic long-term archiving process – where bit preservation addresses storage in an ideally redundant and monitored way, where logical preservation addresses rendering capabilities in the light of general technological change and obsolescence, and where semantic preservation maintains a very general interpretability of the object, e.g., seen in connection with a structured schema contained within the digital object.

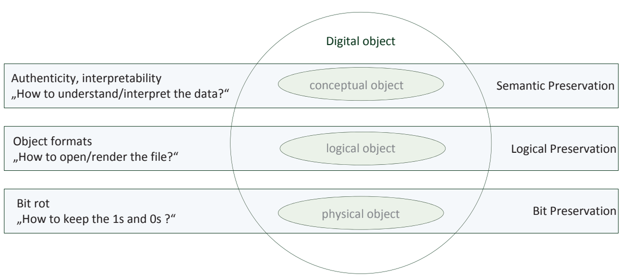

Figure 4. *Three preservation layers of a digital object*

Especially the impact of decisions on the bit preservation and logical preservation layers play a pivotal role in the context of preservation planning. Here, stakeholders' expectations in data storage – for example whether the data should be kept on online, nearline or offline storage – are compared to organizational capabilities and requirements of the OAIS itself, forming a basis for preservation action decisions. The same holds true for the choice of file format or rendering environment. Designated community, technological and organizational factors alike need to be taken into consideration during preservation planning but also in the overarching preservation strategy itself [9].
While the preservation layers model documents the static state of a digital object at any given point in time, the lifecycle model sees the digital object in the constant context of change – may it be change in form of changing ownership or change in form of actions preformed on the object itself. While many domains have created their own lifecycle models - e.g., the DDI combined life cycle model for social, behavioral and economic sciences[6] – the Digital Curation Centre (DCC)[7] has created a generic and high-level curation lifecycle model, which describes lifecycle actions from an object's creation to its use and transformation [11]. The DURAARK project has adapted the DCC curation lifecycle model and created a simplified version for its own use (Figure 5). After the stakeholder identification, interviews were conducted with organizations matching the profiles of the stakeholder groups identified. In the course of these interviews it became clear, that very little awareness of risks associated with digital object decay exists outside of the cultural heritage domain. The simplified object lifecycle model became a helpful tool in communicating the processes associated with curation and preservation practice, as it clearly shows the impact that the fabrication quality of an object may have over the course of time.

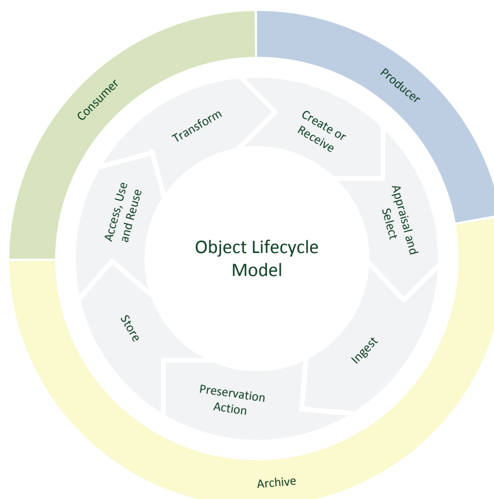

Figure 5. *Digital Object Lifecycle Model (based on DCC Curation Lifecycle Model*

The stakeholder analysis put forth a number of use cases which the DURAARK project aims to address. Some of the use cases specifically target future usage scenarios, such as "Detect differences between planning and as-built state" or "Plan, document and verify retrofitting/energy renovation of buildings". While most of the use cases require tool development as part of the project, for example in form of software to compare pointclouds to BIM objects in order to detect differences between "as-planned and as-built", all use cases put forth requirements for information being passed along from the producer. This may be as simple as the geo location which allows us to match a plan to a scan or as complex as information required for energy renovation, such as the heat transfer coefficient. As no comparable preservation efforts focusing on BIM and pointcloud data have been conducted before, no reference information about data which needs to be included exists. Furthermore, as pointed out in the previous chapter, this information is highly dependant on the stakeholder

[6] `http://www.ddialliance.org/Specification/DDI-Lifecycle/`
[7] `http://www.dcc.ac.uk/`

which either produced the data or queries the data to execute one of the sketched use cases.

Results and Future Work

Two of the main tasks of an OAIS compliant archive in connection to producer and consumer interaction are to ensure the "appropriateness" of data received from the producer and to determine the knowledge that can be assumed of the designated community (see figure 2). The previous chapters identified the stakeholders of 3D architectural data and showed how both, the context in which they created data in as well as the context in which they intend to use data in, has implications on the object itself and on preservation action. The work conducted in the context of the DURAARK project so far shall lay a foundation for a more in-depth look of stakeholder practises and requirements, especially in connection to the use cases formulated within the project.[8] In-line with the two standard digital preservation models introduced earlier (see figures 4 and 5), the further analysis can be divided into two categories:

- Object centric
- Process centric

The object centric analysis shall ensure that the data the OAIS accepts from the producer is appropriate in the light of the data producers' context but also in the light of future usage scenarios. Furthermore, a better understanding of the stakeholders' requirements in an object is essential knowledge in preservation planning activities. From an object centric view, the stakeholders' definitions of the following factors need to be analyzed further:

1. Authenticity
   It is currently unclear how the different stakeholders define authenticity. This question is closely tied to requirements in audit trails and provenance trails, especially in the case of BIM objects containing the information of various producers. Are there, for example, requirements in an audit trail left by a construction company enriching an object with vendor information for parts during the physical construction phase? How do the requirements differ across the different stakeholders?

2. Completeness
   As figure 2 shows, different stakeholders have their own definition of what constitutes "completeness" of data. As completeness is a key factor of appropriate information to be demonstrated by the OAIS, a definition of what "completeness" means needs to be analyzed for each of the targeted stakeholders. The producer's definition of completeness needs to be compared against the requirements implied by the sketched usage scenarios.

3. Quality
   It is to be expected that the stakeholders have different definitions of what constitutes good quality of data. The question of quality is closely tied to authenticity and completeness, but may introduce other factors like behavioral aspects of the data connected to processing requirements.

[8]A full list of the use cases can be found in DURAARK Deliverable D2.2.1 [10]

While the object centric view gives an insight into the data itself – the "what", so to speak, the process centric view sheds light on the "how". An analysis of process centric factors gives insight into how data is currently being handled and archived. It may allow insight into how a sufficient level of control over the data can be achieved and how the stakeholder intends to access the data in the future. From a process centric view, the stakeholders' definitions of the following factors need to be analyzed further:

1. Retention time
   The retention time gives insight into how long the stakeholder currently holds the data available for. Retention times may be based on legal requirements, fixed business regulations or contractual agreements. Besides giving insight into how long the data needs to be available for, retention times indicate organizational or legal dependencies and furthermore hint at future usage scenarios.

2. Usage scenarios
   While the alignment of the stakeholders along the building lifecycle as well as along the object's lifecycle gave a good first insight into the intent with which data will be accessed in the future, usage scenarios should be analyzed in more detail. It is important to note that an analysis of usage scenarios should not only include existing usage processes but also hypothetical or ideal scenarios, which the stakeholder currently sees as not fulfilable due to technical or organizational constraints.

3. Curational practice
   Curational practice analyses how the stakeholder currently processes the information in regards to maintenance – this includes questions about any form of records or document management system as well as requirements the stakeholder may have formulated for third parties who contribute to the data production process, such as for example external companies conducting the 3D scanning.

4. Archival practice
   While the analysis of curational practice is more centered on accompanying information and processes, archival practice shall put forth whether the stakeholder is already aware of any requirements and has for example implemented preservation processes at an organizational level. This is especially of interests for those stakeholders, who have been previously identified as potential archives.

Conclusion

As 3D object processing of architectural data is a multi-faceted process involving numerous stakeholders, the requirements and needs of these stakeholders need to be considered when developing processes and methods for digital curation and preservation of the respective data. The DURAARK project has taken a first step by defining the different stakeholders for architectural 3D data and their potential roles a producer, consumer or archival entity. In connecting stakeholders to preservation requirements, different models need to be considered to match static object requirements with fluent process requirements. The object-centric preservation layer model and the process oriented lifecycle model proved to be helpful tools in the stakeholder alignment. While the alignment was conducted at a high granularity, a number of object-centric and process oriented factors were put forth which shall be analyzed further in the future to improve the understanding of stakeholder requirements and future usage scenarios of the content type.

Acknowledgments

This work was partially funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 600908 (DURAARK) 2013-2016.

References

[1] Stephan Strodl, Petar Petrov, Andreas Rauber. Research on Digital Preservation within projects co-funded by the European Union in the ICT programme. (2011)

[2] Tom Rosko. MIT's FAÇADE Project – Future-Proofing Architectural Computer-Aided Design. CITRA2010, Oslo, Norway. (2010)

[3] Jörg Brunsmann, Wolfgang Wilkes, Gunter Schlageter, Matthias Hemmje. State-of-the-art of Long-term Preservation in Product Lifecycle Management. International Journal on Digital Libraries, 12, pg. 27-39. (2012)

[4] DURAARK. D6.6.1: Current state of 3D object digital preservation and gap-analysis report. Deliverable, DURAARK project. (2014)

[5] DURAARK. D7.1: Current state of 3D object processing in architectural research and practice. Deliverable, DURAARK project. (2014)

[6] David Barber, and Jon Mills. 3D Laser Scanning for Heritage – Advice and guidance to users on laser scanning in archaeology and architecture, School of Civil Engineering and Geoscences, University of Newcastle. (2011)

[7] Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS). (2012)

[8] Kenneth Thibodeau. Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. In: The State of Digital Preservation: An International Perspective. CLIR Report 107. (2002)

[9] Michelle Lindlar. Time to Change: Effects and Implications of Digital Preservation in an Organizational Context. IASSIST Quarterly, Issue 3-4, pg. 8-12. (2012)

[10] DURAARK. D2.2.1: Requirements Document. Deliverable, DURAARK project. (2013)

[11] Sarah Higgins. The DCC Curation Lifecycle Model. The International Journal of Digital Curation, Issue 3, pg.134-140. (2008)

Author Biography

Michelle Lindlar is Technical Analyst and Digital Preservation Researcher at the German National Library of Science and Technology (TIB) in Hannover, Germany. Before joining TIB she worked at the German National Library of Medicine (ZB MED) as a digital preservation project manager. She leads the work package on long-term preservation within the DURAARK project.

Martin Tamke is Associate Professor at the Centre for Information Technology and Architecture (CITA) at the Royal Academy of Fine Arts, School of Architecture in Copenhagen where he is pursuing a design led research on the interface and implications of computational design and its materialization. Within the DURAARK project he leads the work package on data acquisition, evaluation and testing.