

MonArch – Digital Archives for Monumental Buildings

Burkhard Freitag, Christoph Schlieder

Digital archives form a technological basis for preserving our cultural heritage. This article focuses on archives for monumental buildings and discusses what semantical support is needed to achieve both usability and interoperability.

1 Introduction

Monumental buildings such as cathedrals or castles together with their often very old and large archives form an important part of our cultural memory and heritage. Unique documents like medieval manuscripts, precious incunables, and church registers as well as original building plans, construction reports, old chemical recipes used for repairing and various others can be found in these archives. However, in many cases only a fraction of the existing documents have been catalogued, inventoried, digitally secured and preserved. Moreover, it is not straightforward to relate documents to each other based on their contents or properties.

Even if the archived documents have been digitized and are digitally stored, which is the aim of many research projects all over the world, retrieval based on structural neighborhood or similarity, for example asking the query "give me all documents related to a certain façade segment", is only scarcely supported by existing digital archives for monumental buildings. Mostly, the physical documents are merely stored in chronological order in a more or less systematic way. There is practically no index or register relating the documents to the building's structure.

Similarly, it is close to impossible to systematically exploit the document base of a monumental building based on properties like the kind of damage occurring in structural parts of the building, given a certain architectural category or cultural style. The only way to find the desired information is to sequentially scan the entire archive which often means having to inspect thousands of physical documents.

Presently, most archives of monumental buildings are isolated in the sense that they serve as a stand-alone store of physical or – in some cases – digital documents. Therefore, it is also impossible to interconnect different archives as postulated by Borgman [2]. As a consequence, semantically connected documents that are spread over different archives cannot be related to each other.

Another issue arises from the missing standard of document descriptions. To date no consistent set of metadata and no metadata model exist, either for the structural model of the building or for other descriptive categories such as material used, kind of damage observed, architectural category, cultural style. Therefore, asking queries across different archives and buildings is almost impossible, let alone combining several digital archives in a peer-to-peer network as proposed in recent research (see e.g. [7]).

In this paper, the *MonArch* framework [15] for distributed web-based digital archives for monumental buildings is described¹. The *MonArch* project can be seen as a contribution to built heritage preservation, i.e., keeping monumental buildings in good physical order, understanding their cultural and historical context, and conserving all related documentation. The digital archives considered are centered around construction- and maintenance-related documents, but also can and do accommodate documents mainly describing artifacts like paintings, sculptures and others. The *MonArch* framework supports spatial and structure-oriented queries, context-based search and retrieval, as well as an extensible indexing scheme based on a multidimensional metadata model. The system is cooperation-ready by both its communication architecture and its ability to use ontologies for semantical information integration. The *MonArch* system is already in operational use, but is nevertheless evaluated constantly and adapted to real life needs by a project group consisting of master builders of various monuments, restoration scientists, architects, and computer scientists. Currently, the evaluation is mainly performed from a technical point of view and monitors parameters such as general usability, functional completeness, adaptability and performance.

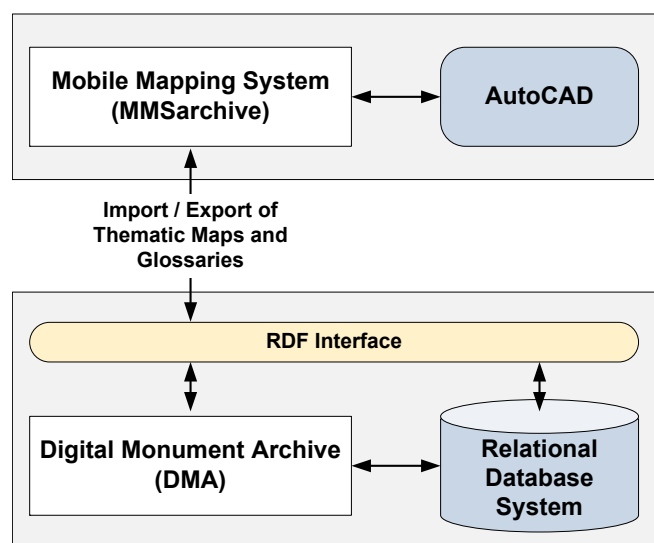


Figure 1: Main Components: MMSArchive and DMA

¹ This work is funded by the German Research Foundation (DFG) under contracts DR 490/1-1, FR 1012/8-1 and SCHL 447/5-1

2 Overall Architecture

The *MonArch* framework consists of two main components (see Figure 1): *MMSarchive* is a system supporting the semantics-guided construction of graphical maps and the thematic mapping of various phenomena related to buildings. The *Digital Monument Archive (DMA)* allows the storage and retrieval of digital documents according to their metadata and by navigation along the structure of the building under consideration. Both components can be coupled to exchange their thematic maps and glossaries based on the standard graphics format DXF [1] and on RDF data [16].

The DMA component has a client-server architecture (see Figure 2). A role-based access control and a transaction management ensure that user access to the archive is isolated; this applies in particular to updates of metadata that is normative for the entire archive such as glossaries and structural metadata.

MonArch DMA servers can be connected over the internet using web services. At the technical level this allows the submission of distributed queries spanning several archives. Even external non-*MonArch* archive servers can be included. The "Bildarchiv Foto Marburg" (<http://www.fotomarburg.de>), for instance, is currently in the process of being connected to the *MonArch* system using the standard *museumdat* [5].

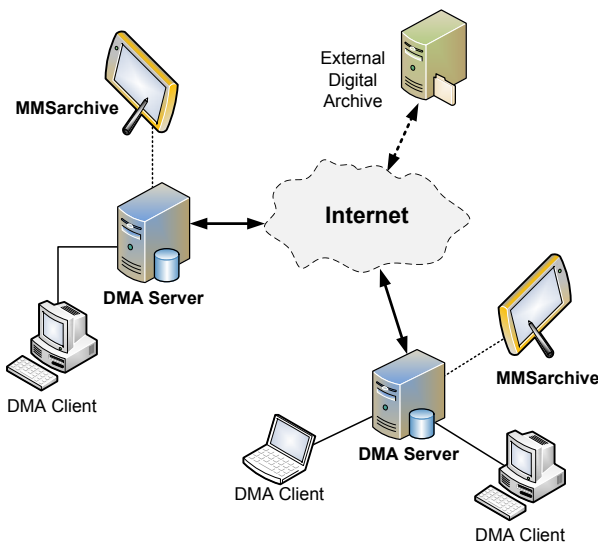


Figure 2: *MonArch* Network

3 Digital Maps

Among the documents stored in the DMA, the architectural drawings or maps that are produced with the *Mobile Mapping System* in its archiving edition (*MMSarchive*) play a special role. Workflows based on hand drawn maps have dominated built heritage preservation until recently [14]. Digital documentation emerged only a decade ago, much later than in architecture. In part, this technological delay can be explained by the specific mobility requirements of preservation science. Architects work from the design towards the artifact while preservation scientists generally proceed in the opposite direction. They solve a reverse engineering problem

which consists of trying to understand the design principles behind an undocumented historical building. Whereas architects need computational resources in their office, preservation scientists need them on the scaffold at the construction site. It was not until inexpensive and ruggedized mobile hardware appeared on the market that digital documentation became feasible in built heritage preservation ([10], [4]).

Preservation scientists who document a building have to address a number of subtasks: modelling the building's structure, classifying the damages, planning the preservation methods, and recording the methods that have been applied. Each of the subtasks produces a digital map as its result document. Two important types of maps are:

- Inventory maps that describe the different parts of a building with their location and dimensions as they are determined by more or less complex measurement procedures, e.g. tachymetry, photogrammetry, or laser scan.
- Damage maps that are produced by domain experts who inspect the damaged parts of the building and analyze the type, the severity and the causes of the damages.

Digital maps are rather complex digital documents which associate geometric and thematic data. Note that map detail varies very much. Some buildings of great historical significance have a permanent workforce devoted to preservation which produces inventory maps representing every ashlar, that is, every stone that has been worked by a mason. Figure 3 shows a damage map which has been produced by the lodge of Passau cathedral. Point, line and polygon features are arranged on several layers. The background layer is formed by a raster image, a historical inventory map. A second layer depicts the basic components of the construction, ashlars and joints. It serves as a vector representation of the inventory map. The remaining layers are used to represent different types of damage and constitute the damage map.

Each geometrical object from the inventory or damage map is associated with thematic data. When a preservation scientist inspects part of a building, he or she documents the spatial extent of a damage by drawing with an active pen on the screen and by entering data associated with the damage into a form similar to the one shown either by handwriting or using a screen keyboard. Different data forms are associated with different types of damages and are generated automatically from the data model. Since no two buildings are exactly the same and since the preservation methods adopted differ from case to case, a flexible approach to data modelling that includes means for specifying the semantics of the thematic data constitute a central requirement of our user group.

So far, two off-the-shelf technologies have been used for documentation purposes in built heritage: AEC (architecture, engineering, construction) solutions like AutoCAD and GIS (geographic information systems) such as ArcView [10]. Both, AEC and GIS technologies, consist of spatial databases that provide some flexibility of the thematic data models but do not support the explicit modelling of data semantics. *MMSarchive* follows an ontological engineering approach by modelling the semantics of thematic data in OWL-DL [17]. For the drawing functionality a standard CAD is used which also acts as a spatial data repository: *MMSarchive* is implemented as an AutoCAD application.

Ontologies have been studied in connection with cultural heritage in general. Most efforts went into designing appropriate top-level ontologies like the CIDOC conceptual refer-

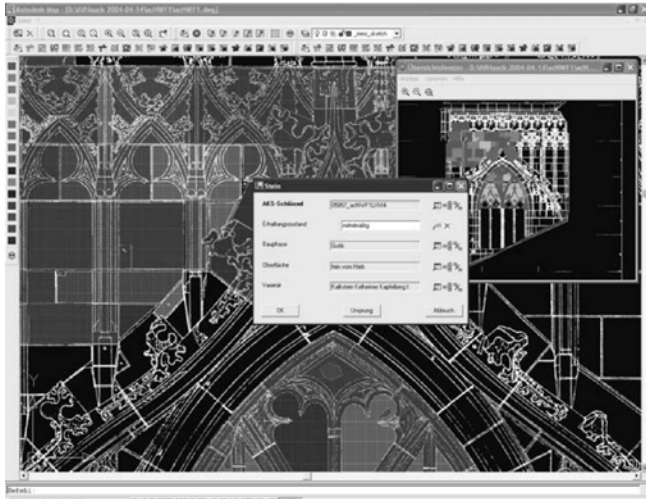


Figure 3: Damage map produced with the MMSArchive

ence model that is mainly used by museums [3]. The specific issues relevant to the built heritage domain, that is, a domain ontology grounded in spatial objects, have not been addressed. A similar type of problem, however, has been studied in GI science in connection with semantic interoperability [9].

4 Ontological Reasoning Support

An important functionality of MMSArchive consists in automatically generating the metadata that describes the content of a digital map. In interacting with the documentation system, the user provides valuable semantic information while producing the map. Through the drawing tool, every geometrical object created is automatically associated with information about its categorization, i.e. the class of the domain ontology it belongs to. In addition, the user specifies properties and associations for most of the spatial objects using a domain ontology of his or her choice. The interpretation of properties and associations with respect to the domain ontology is transparent for MMSArchive.

In other words, the challenge of metadata generation consists not so much in extracting the semantics but in choosing which part of the semantic information should be described in the metadata. Users almost never need to formulate queries that involve a specific geometric shape. The shapes of stone damages are highly irregular and they do not possess any diagnostic value. Thus, it is not surprising that our test users did not feel the need to query for a crumbling with hexagonal shape. We therefore chose to generate metadata by adapting the approach of thematic projection described by [13]. It consists (1) in completely abstracting from all geometrical shape information, and (2) in partly abstracting from information about position. Only qualitative information about the position is retained by encoding topological relations that hold between objects of the damage map and spatial objects that serve as a spatial reference system. Typically, the objects of the inventory map are used for that purpose, in which case the thematic projection of a digital map describes how the thematic objects of the damage layer relate to the ashlar of the inventory map. Spatial rela-

tionships are expressed using the RCC-5 system of topological relations. It is this information that is used as metadata and encoded in an ontology (OWL TBox) and a knowledge base (OWL ABox). The spatial relationships resulting from the thematic projection are expressed explicitly as relationship assertions.

Thematic projection only represents a subset of the possible spatial relationships between objects. Yet, this reduced set of relationships is sufficient to enable certain consistency checks at the ontological level as for example has also been shown by [8].

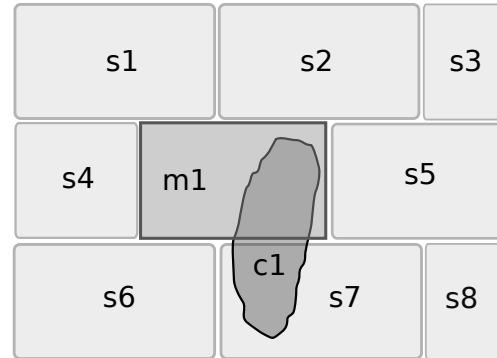


Figure 4: Example Situation for Spatial Consistency Checking
Ashlars (light grey) and MetalPlatings (medium grey) are spatial reference objects.

Consider for example the schematic representation of a mapping in Figure 4. The shaded areas s_1 to s_8 represent ashlars, while the dark area m_1 in the center is a metal plating. The overlapping transparent object c_1 describes a crumbling damage to the underlying ashlars. We assume that ashlars and the metal plating are spatial reference objects.

After thematic projection, two relationship assertions are inserted into our exported knowledge base: $overlaps(c_1, m_1)$ and $overlaps(c_1, s_7)$. If we now interpret the $overlaps$ -relation for crumbling as a reference to the affected objects, a problem occurs: the map represents a crumbling damage on a metal plating. Yet, from domain knowledge it is known that crumbling damages only occur on ashlars. The depicted situation is logically inconsistent.

This consistency condition can easily be axiomatized in description logics. Equations 1 to 4 show the described situation. A theorem prover is used to detect the inconsistency.

$$MetalPlating \sqsubseteq \neg Ashlar \tag{1}$$

$$Crumbling \sqsubseteq \forall overlaps. Ashlar \tag{2}$$

$$Crumbling(c_1), MetalPlating(m_1) \tag{3}$$

$$overlaps(c_1, m_0) \tag{4}$$

The explicit modelling of data semantics also permits to address another issue: format migration, which is linked to the problem of long term archiving. There is common agreement that the interpretation barrier [11] forms the most significant unsolved problem with regard to the long term preservation of digital documents. The situation is made

even more severe by the widespread use of proprietary, often undocumented data formats. The choice of open, documented data formats for the representation of digital maps is thus expected to simplify their preservation for the long term.

Yet, open data formats do not solve all problems of data preservation. Data formats change over time as the representation of information is adapted to changing external needs. This format evolution process makes it necessary to migrate data from outdated formats. The use of formal ontologies with their well defined semantics enables us to map the format migration problem into the domain of schema and ontology mapping [6]

5 Spatial Queries and Multidimensional Querying

During requirements analysis we observed architects, maintenance experts and conservation specialists doing their work and found that they tend to locate documentation in the same way they read plans or explore “their” monument: by following its architectural structure. For this reason, some non-digital archives provide a kind of a spatial index which, however, may be just as simple as a list of the major parts of a cathedral. Due to a “structural binding” of archival items it is easy for the digital archive to support this kind of query, even in a rather elaborate way if needed. Any item stored in the archive can be assigned to a structural element, i.e., to a composite such as a façade segment or to an atomic element like a single stone. The hierarchical structural decomposition of the building is represented by the underlying *partonomy* (cf. Figure 5). There are additional descriptive categories, i.e. *topics*, an item can be related to. As an example, consider a category describing the kind of damage observed, or another category telling which cultural style a specific item belongs to. In general, the descriptive categories are structured as well. As an example consider a category that captures the kind of damage possibly found in a particular part of the building. It is worth mentioning that multiple assignments are possible, i.e., an archival item can be assigned to more than one structural element and descriptive category. The idea of a more sophisticated spatial and descriptive index was well received by our test users right from the beginning and proved to be a highly useful functionality as the system evolved.

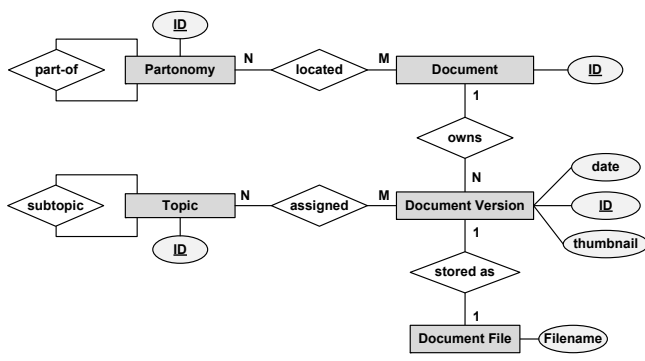


Figure 5: Simplified DMA Data Model

For convenience, the tree-like structural view on the archive is supplemented by an interactive map showing a (simplified) map of the entire building or some of its parts (see Figure 6). The map is tightly attached to and synchronized with the tree-view. This allows to use the map as an alternative way to specify the structural part of the building to be inspected just by clicking. Vice versa, the parts of the building selected via the tree-view are highlighted to give the user an intuitive orientation.

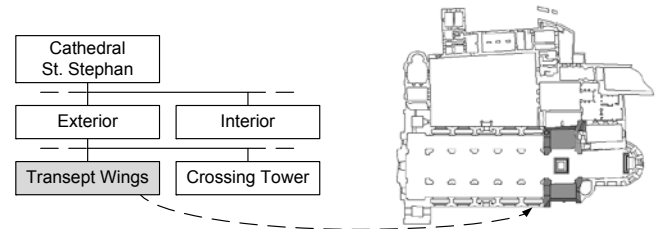


Figure 6: Synchronizing Partonomy and Map in DMA

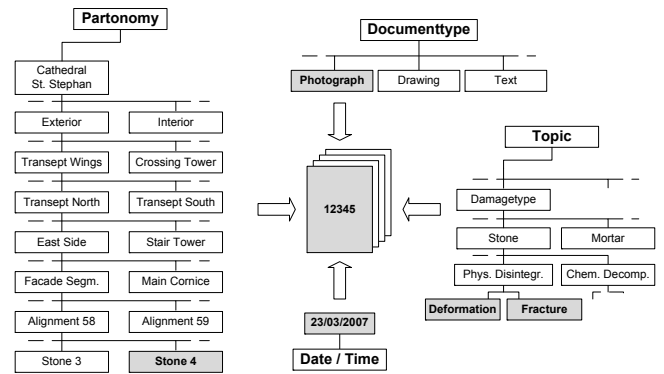


Figure 7: Multidimensional Querying

From the user’s point of view, the multidimensional representation and querying of archival items can be interpreted as in the example shown in Figure 7. There, the document with identification number 12345 is a photograph showing stone 4 of alignment 58 of the façade segment on the east side of the northern transept, which is part of the transept wings seen from the exterior of the cathedral St. Stephan. It is known that this photograph, which has been taken on 23 March 2007, shows deformation or fracture of stone 4. Note that more than one damage type but only a single structural element (stone 4) have been attached to the sample document. Posing a more general query, we could have asked for photographs showing some kind of physical disintegration on the northern transept or one of its subparts. Intuitively, one would expect document 12345 to be contained in the corresponding answer set, too.

As the example illustrates, structural queries are not restricted to atomic structural elements. The semantics of a structural query can therefore be *pointwise*, i.e., only those items attached to the specified structure element are retrieved, or *including subtrees*, i.e., the items attached to the particular element specified or to any of its parts form the answer set, or *including ancestors*, i.e., all items are returned that are attached to the element specified or to any of the structural elements it is a part of. Analogously, queries along

the descriptive categories can be pointwise, including the subtrees, or including the ancestors.

6 Interoperability

One of the big issues in digital preservation of cultural heritage is interoperability. Multiple users tend to define their own metadata vocabulary resulting in incompatible assignments of archival items. Distributed digital archives, though connected via the internet on a technical level, cannot cooperate on a semantical level since their semantical categories do not coincide. Of course, one can try to solve part of the problem in a more or less administrative way by applying *authority control*. The *MonArch* components incorporate some standard ontologies such as CIDOC CRM [3] and SWD [12] that contain standard descriptors arranged in a generalization hierarchy.

When considering authority files as defining just the upper part of an applicable ontology, individual users or local archives can define subcategories of descriptive terms rather freely. Queries can be restricted to the normative descriptors as defined by the authority file. Provided query answering is run in subsumption mode, i.e. include subtrees (see Figure 8, upper part), the result is still compatible with other archives but includes archival items assigned to the more fine-grained descriptors. It is well known that this kind of query is supported by one of the essential inference services in the field of ontology-based reasoning. As long as every node of a *MonArch* network sticks to the normative metadata, even queries spanning more than one digital archive can be accommodated. The answer set is simply computed by running the same query on every local archive and combining the partial answers using set union (see Figure 8). Similarly, query answers can be restricted to only those documents having descriptors common to all archives accessed.

Elaborate authority files do not only contain the normative metadata themselves, but also meta-metadata, i.e., descriptors for metadata. In this case, information about synonyms can be taken into account as a possible next step, i.e., the answer set can be extended by following the synonym relationship. If more information is available, ontology-based reasoning can go even further by applying similarity metrics on the set of descriptors for query answering. In this case, also documents having descriptors "similar" to those used in the query will be retrieved.

As an essential prerequisite for coupling the two main components, i.e., the MMSArchive and the Digital Monument Archive, RDF has been chosen as the exchange format (see Figure 1). Since the Digital Monument Archive is basically built on top of a relational database system and all information is stored as entries in relations, RDF data is transformed into relational data query and vice versa at import / export time. Consequently, ontology-based reasoning either has to be transformed into relational query answering or has to be applied in a preprocessing step to provide for the appropriate relational indexing. The DMA supports both approaches.

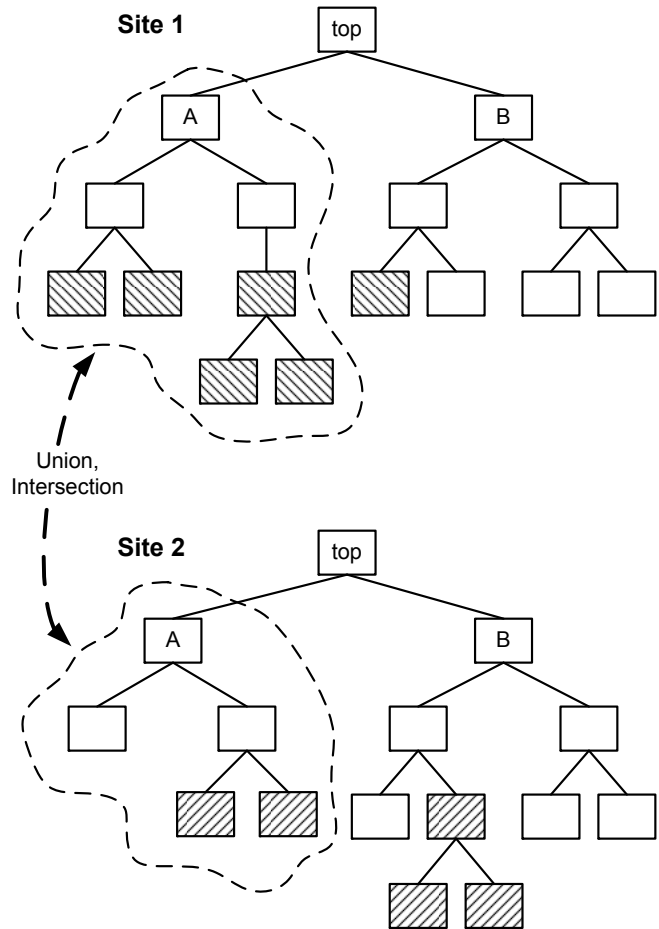


Figure 8: Using Authority Files for Semantical Interoperability

7 Summary

Digital archives have been introduced that address two major functionalities needed for preserving monumental buildings and their valuable documentation: map creation with semantically guided thematic mapping and document management supporting spatial access and multidimensional querying. It has been shown that interesting research questions arise in the intersection of semantic modeling and intelligent retrieval of documents, some of them involving ontological reasoning, others addressing support for efficient retrieval and interoperability.

References

- [1] Autodesk Inc. *DXF Reference*, 2008. available at http://images.autodesk.com/adsk/files/acad_dxf.pdf.
- [2] C. L. Borgman. Challenges in Building Digital Libraries for the 21st century. In *ICADL '02*, pages 1–13, 2002.
- [3] M. Doerr, C.-E. Ore, and S. Stead. The CIDOC conceptual reference model: a new standard for knowledge sharing. In *International Conference on Conceptual Modeling*, pages 51–56. Australian Computer Society, Inc., 2007.
- [4] R. Eppich and A. Chabbi. *Recording, Documentation, and Information Management for the Conservation of Heritage Places: Volume II: Illustrated Examples*. Getty Conservation Institute, 2007.
- [5] R. Stein et al. museumdat - XML Schema zur Bereitstellung von Kerndaten in museumsübergreifenden Beständen. Technical report, Fachgruppe Dokumentation im Deutschen Museumsbund

/ Institut für Museumsforschung SMB-PK / Zuse-Institut Berlin, 2007. available at <http://museum.zib.de/museumdat/museumdat-v1.0.pdf>.

- [6] J. Euzenat and P. Shvaiko. *Ontology Matching*. Springer, 2007.
- [7] Ingo Frommholz, Predrag Knezevic, Bhaskar Mehta, Claudia Niederée, Thomas Risse, and Ulrich Thiel. Supporting Information Access in Next Generation Digital Library Architectures. In *Digital Library Architectures: Peer-to-Peer, Grid, and Service-Oriented, Pre-proceedings of the Sixth Thematic Workshop of the EU Network of Excellence DELOS, S. Margherita di Pula, Cagliari, Italy, 24-25 June, 2004*, pages 49–60. Edizioni Libreria Progetto, Padova, 2004.
- [8] T. Hahmann and M. Gruninger. Detecting Physical Defects: A Practical 2D-Study of Cracks and Holes. *AAAI Spring Symposium*, 2009, 2009.
- [9] W. Kuhn, M. Raubal, and P. Gärdenfors. Cognitive Semantics and Spatio-Temporal Ontologies. *Spatial Cognition and Computation*, 7(1):3–11, 2007.
- [10] R. Letellier, W. Schmid, and F. LeBlanc. *Recording, Documentation, and Information Management for the Conservation of Heritage Places, Volume I: Guiding Principles*. Getty Conservation Institute, 2007.
- [11] R.A. Lorie. Long term preservation of digital information. *Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries*, pages 346–352, 2001.
- [12] Deutsche Nationalbibliothek. Schlagwortnormdatei (SWD). available at <http://www.d-nb.de/standardisierung/normdateien/swd.htm>.
- [13] C. Schlieder and T. Vögele. Indexing and Browsing Digital Maps with Intelligent Thumbnails. In *Tenth International Symposium on Spatial Data Handling*, pages 781–782. Springer, 2007.
- [14] M. Schuller. *Building Archaeology, Monuments and Sites: VII*. International Council on Monuments and Sites (ICOMOS), 2002.
- [15] MonArch Team. The MonArch Project. <http://www.monarch-project.eu>.
- [16] W3C. Resource Description Framework (RDF) Model and Syntax Specification. <http://www.w3.org/TR/PR-rdf-syntax/>, 1999.
- [17] P. Wullinger and C. Schlieder. Digital maps of historical buildings: Preservation issues and solutions. In *Proceedings of IS&T's Archiving 2008 Conference*, June 2008.

Contact

Prof. Dr. Burkhard Freitag
Lehrstuhl für Informationsmanagement
Universität Passau
94030 Passau
Burkhard.Freitag@uni-passau.de

Prof. Dr. Christoph Schlieder
Lehrstuhl für Angewandte Informatik
in den Kultur-, Geschichts- und Geowissenschaften
Universität Bamberg
Feldkirchenstrasse 21
96052 Bamberg
christoph.schlieder@uni-bamberg.de



Burkhard Freitag is a Professor of Computer Science at the University of Passau. His research activities are in the areas of databases and information systems. Currently, his work concentrates on digital archives, document verification and context-sensitive information systems using methods from formal logic, database systems and knowledge representation.



Christoph Schlieder holds a doctoral degree as well as a habilitation degree in Computer Science from the University of Hamburg. Since 2002, he teaches Computing in the Cultural Sciences at the University of Bamberg. His research focuses applying methods from ontological engineering and from qualitative spatial reasoning to problems that arise in the cultural sciences.