MonArch – A Digital Archive for Cultural Heritage

BURKHARD FREITAG AND ALEXANDER STENZER

DEUTSCHSPRACHIGE ZUSAMMENFASSUNG

MonArch – Ein digitales Archiv für das kulturelle Erbe

Moderne Technologien erlauben nicht nur die Erstellung präziser digitaler Modelle von Denkmalen, sondern versetzen uns auch in die Lage, sehr große Bestände an Dokumenten, Inkunabeln, Originalplanzeichnungen, Fotografien, Artikeln, Büchern und anderen Archivalien zu digitalisieren, die mit dem Monument verbunden sind. In vielen Fällen allerdings sind die physischen Dokumente über verschiedene Orte verteilt, mehr oder weniger unsystematisch abgelegt, nur teilweise katalogisiert, und eine semantische Auszeichnung oder Verschlagwortung existiert oftmals nur in unzureichender Form. Unglücklicherweise überträgt sich diese Situation nicht selten auf die erzeugten digitalen Versionen der wertvollen Archivalien. Dies ist teilweise der großen Masse an erzeugten Informationen geschuldet, teilweise aber auch der Tatsache, dass neue Organisationsformen für den digitalen Informationsbestand erforderlich sind, um sowohl seine breite Nutzung als auch seine Bewahrung sicherstellen zu können.

Das MonArch System, das wir in diesem Beitrag beschreiben, bietet Unterstützung für die nachhaltige Speicherung und das Wiederauffinden von digitaler Information und Dokumenten auf der Grundlage ihrer Beziehung zu einer räumlichen Struktur. Für eine solche Organisationsform digitaler Information besonders geeignete Strukturen sind beispielsweise Bauwerke und urbane Situationen sowie archäologische Stätten.

SUMMARY

Modern technologies not only enable us to create precise digital models of historic monuments, but also allow us to digitize the enormous collections of documents, incunabula, original plans and drawings, photographs, articles, books and other archival materials that are associated with these monuments. In many cases, however, the physical documents are distributed among different locations, are more or less unsystematically filed and are only partially catalogued, and any semantic tagging or keyword indexing that might exist is often insufficient. Unfortunately, this is also frequently the case with the digitized versions of valuable archival materials. The problem lies in part with the large volume of data being generated, but also with the fact that new forms of organization for digital information collections are needed in order to ensure both their broad usability and their preservation.

The MonArch system described in this essay offers support for the sustainable storage and retrieval of digital information and documents on the basis of their relationship to a spatial structure. Examples of structures especially well served by such forms of digital information organization include buildings and urban situations, as well as archaeological sites.

Introduction

MonArch is an information system designed for documenting structures such as architectural objects, urban situations, and archaeological sites. A MonArch database consists of a digital model of the structure, i.e. a digital representation of the building, ensemble or site, together with a (potentially huge) body of information and digital documents. Any digital or digitized information or document can be attached to the digital model or specific parts of it. Thus MonArch provides a space-related organization of information, documents and artefacts. As an example, consider the query ‘Show all documents related to a certain façade segment of a specific building’.

Of course, there are more descriptive dimensions than just the structural context of a document. MonArch uses vocabularies to assign semantical descriptions such as material used, kind of damage observed, architectural category, or cultural style to arbitrary structural elements as well as to any information or document stored. The sample query ‘Show all documents describing damages of those parts of the building that are made of a particular kind of stone' illustrates the use of additional semantical descriptions.

The MonArch information system has been the outcome of several research projects[1] at the University of Passau, Germany, where software development and maintenance as well as user assistance are still continuing.

The MonArch approach has already acquired a good reputation in the fields of historic monument preservation, archaeology, art history and other areas in Germany and Europe.[2] This technology has been applied successfully at many cathedrals and churches, secular buildings, and archaeological sites, among them well-known UNESCO World Heritage sites. Examples are the Dom St. Stephan (St. Stephen's Cathedral) in Passau (Germany), the churches of St. Sebaldus and St. Lorenz (St. Lawrence) in Nuremberg (Germany)[3], the Old City of Bukhara (Uzbekistan)[4], and the Roman Imperial Baths in Trier (Germany)[5], to name but a few. MonArch is also being used to conduct the Palace Research project in Aachen (Germany)[6].

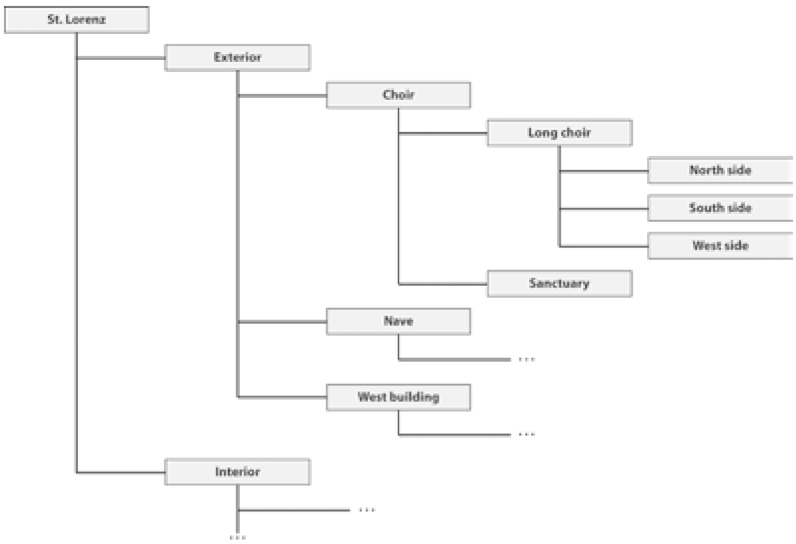

Fig. 2: Part of structural graph of church of St. Lorenz in Nuremberg

Explaining the MonArch approach

The MonArch information model is based on a combination of a graphical visualization (Fig. 1, center window) and a graph representing the structural decomposition of the building or site (Fig. 1, left-hand window). The top node of the graph corresponds to the entire building or site whereas the nodes at lower levels are associated with its structural subparts. As an example, consider the structural graph of the church of St. Lorenz in Nuremberg (Fig. 2) where the entire building consists of two parts, i.e. its exterior and its interior. The long choir – one level below – has the structural subparts north side, south side, and west side. Structural associations such as those described are persistent, i.e. they remain valid and are available in future sessions.

It is worth mentioning that multiple graphical visualizations can be associated with the structural graph depending on the level of granularity desired. For instance, it may be adequate to associate an overview floor plan to the top-level node representing the entire building, but switch to a more detailed plan of a single room, which exhibits some particularly important features. We will later see an example (Fig. 8).

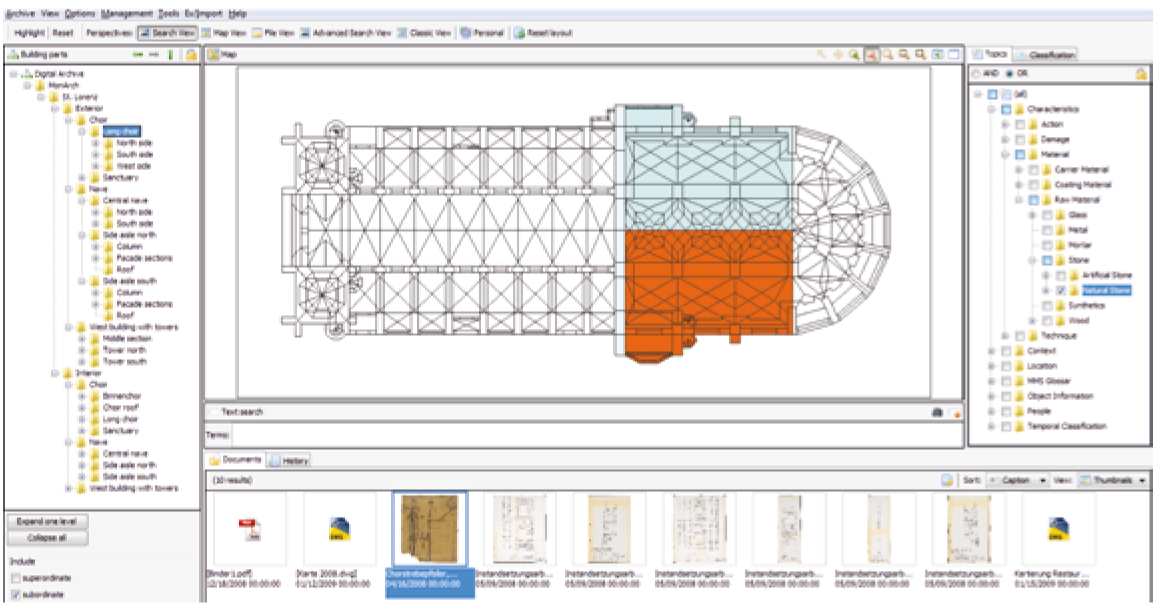

Fig. 1: Screenshot of MonArch representation of church of St. Lorenz in Nuremberg

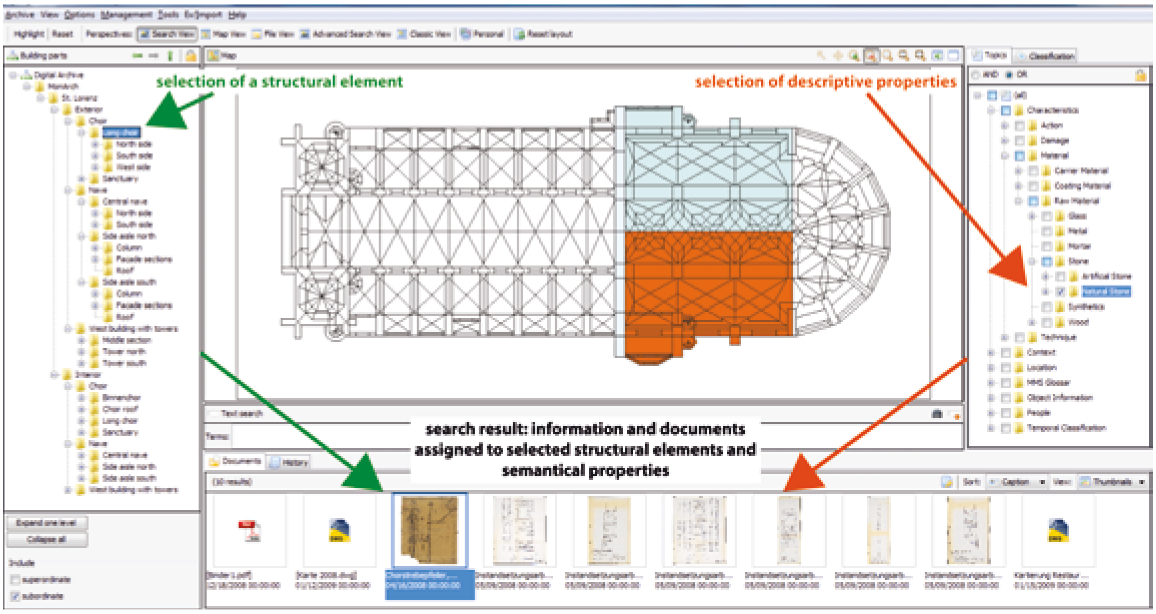

Fig. 3: Retrieving documents by selecting a structural element and additional descriptive properties

Documents and other digital objects stored in the system can be assigned to suitable structural elements, i.e. parts of the building or site. Thus, the hierarchical graphical-structural representation serves as an organizational structure (index) for all information items stored in the system. In addition to the structural representation just described, the information model provides one or more vocabularies (Fig. 1, right-hand window) that can be used for assigning semantical descriptions to structural elements, information items and documents.

When selecting a structural element by mouse-click either in the structural graph (Fig. 3, left-hand window) or in the graphical visualization (Fig. 3, center window), the context of document and information retrieval is set accordingly. Therefore, only the information and documents that have formerly been assigned to the selected structural element are shown in the result window (Fig. 3, bottom window).

In addition to structural selection, semantical properties can be specified by selecting one or more keywords from a vocabulary (Fig. 3, right-hand window), again by a mouse-click. Both the structural context defined and the semantical descriptions provided by the specified keywords form a search filter that restricts the set of information items and documents found.

Sometimes it is convenient to be able to perform a Google-like full-text search, which is also possible (Fig. 3, small window above the result window).

Let us now have a look at a screenshot of a documentation of the Barbara Baths in Trier, Germany, based on MonArch (Fig. 4). By selecting room 18-o in the structural representation (left-hand window), the focus of document retrieval has been set to only those documents that have been assigned to this room. In our example, 68 documents have been found and are displayed (bottom window). The project team working on this application defined their own vocabulary, which they used to attach additional semantical properties to the documents stored.

The next screenshot (Fig. 5) demonstrates the effect of additionally specifying the keyword 'Technique', thus expressing that the search should focus on documents bearing some information about particular techniques applied to the selected room. In the example, the keyword 'Technique' refers to water technique. The set of documents found has now become significantly smaller with only three documents remaining. Specifying additional descriptive properties works like a filter, narrowing down the search for documents and information.

When moving one level down in the structural graph, the visualization window may display a more detailed image depending on the choices made when setting up the representation of the site. In our example, we now select the floor of room 18-o and can observe that the center window has switched to a more detailed plan (Fig. 6) covering just the floor.

The system is able to distinguish and selectively display various graphical/semantical layers. In the example, the first layer shows those parts of the floor of room 18-o that belong to the original Roman phase

(Fig. 7, parts marked red), while the second layer shows the parts of the floor of room 18-o that were restored during the years 1956 – 1962 (Fig. 7, parts marked blue). The layer window indicates that there is a third layer displaying the parts of the floor of room 18-o that were restored during the years 1967 – 1972. Note that the layers as provided here are not mere graphical layers, but can be and are frequently used as 'semantical layers' grouping structural elements that have certain properties in common.

When selecting wall d, a detailed map displaying damaged parts is shown in the center window (Fig. 8). This particular graphic has been deliberately associated with the structural element according to the overall purpose of this specific MonArch application.

Establishing a MonArch database

As we have seen, a MonArch database consists of three major components (Fig. 9). The first component contains the representation of the spatial structure of the object and consists of a structural decomposition in the form of a structural graph and one or more graphical visualizations associated to the structural graph.

The second component consists of one or more concept hierarchies (ontologies) representing additional semantical properties that can be used to describe the objects, information and documents stored. The entire body of data, digital documents and other objects stored in the MonArch database forms the third component.

Defining the structural representation, i.e. the first component, comprises a sequence of mandatory steps that must be followed for each object considered (Fig. 10). Let us first discuss the creation of the graphical visualization. Note, however, that the creation of the graphical model and the definition of the structural graph are essentially independent of each other and can often be performed in parallel.

As a first step, a surveying technique such as manual measurement, laser scanning, or photographing is applied to obtain raw data as a basis for the creation of the graphical model. In a second preparatory step, an appropriate conceptual segmentation of the object must be defined. This is particularly important since the strength of the MonArch approach relies largely on a suitable subdivision of the entire object into semantically sensible parts. A proper segmentation is fundamental to the ability to select a specific spatial context for assigning or

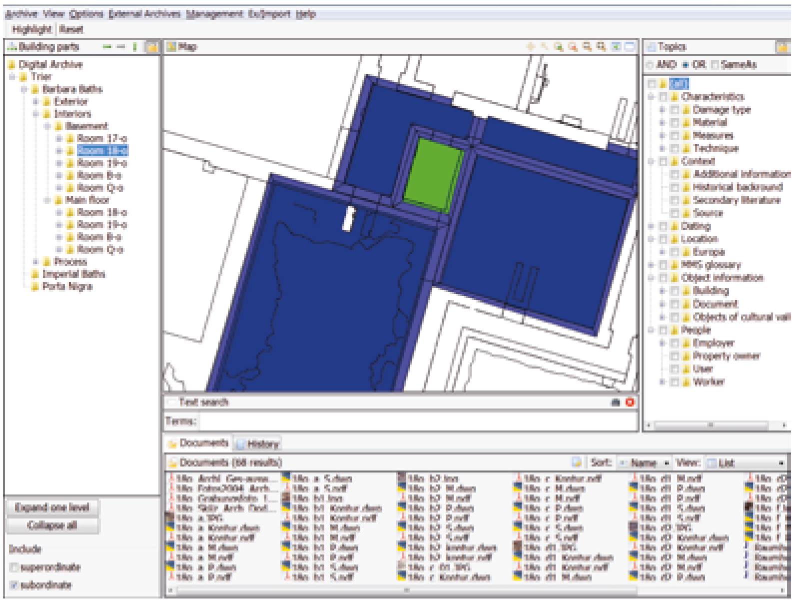

Fig. 4: Barbara Baths in Trier – Select room 18-o

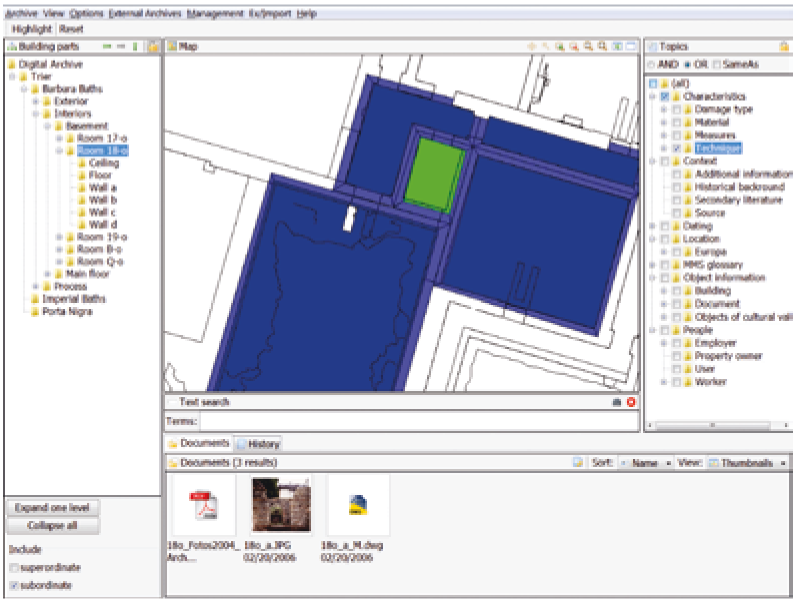

Fig. 5: Specification of additional properties narrows the focus of document retrieval

querying the information stored in the system. Of course, occasionally there may be cases when a reasonable decomposition strategy seems not to exist, forcing the user to assign the entire object as the (only) spatial context.

After the two initial steps, the creation of the graphical model and the definition of the structural graph can be performed more or less independently of one other. A graphical editor or CAD system is used to draw a 2D plan or construct a 3D model based on the results of the measurements, thereby taking the chosen segmentation into account. The resulting graphical model can be imported into the internal database automatically at a later point (Fig.

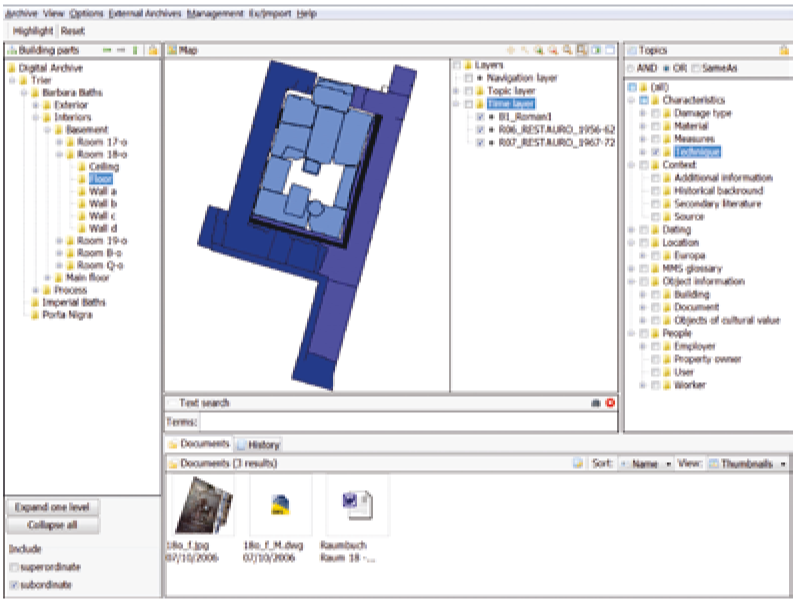

Fig. 6: Barbara Baths in Trier – Selecting the floor of room 18-o causes switch to more detailed floor plan

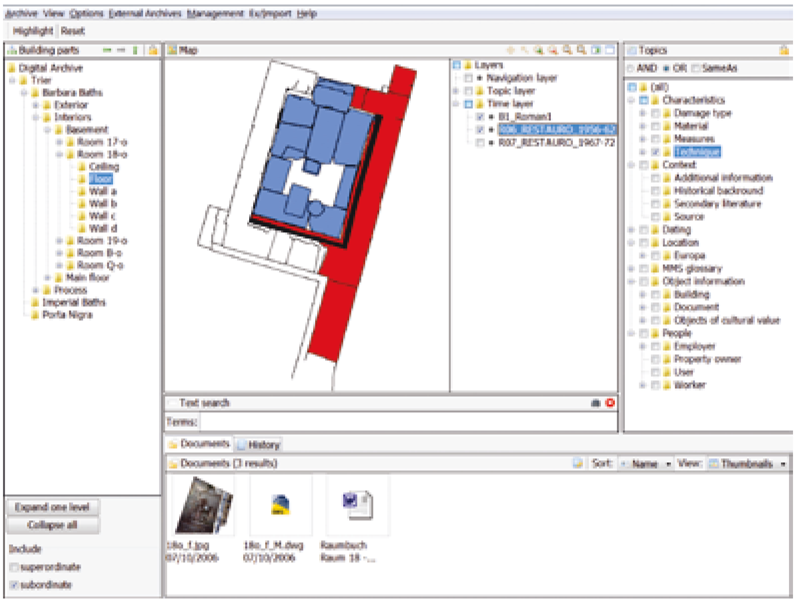

Fig. 7: Floor of room 18-o – Roman phase (marked red) and first restoration phase (marked blue)

10, upper path). The structural graph is defined by adding nodes to the hierarchy, proceeding from top to bottom and starting with the top node representing the entire object. During this process, the resulting graph is automatically stored in the internal database (Fig. 10, lower path). The graphical and structural representations are associated to each other by dragging a graphical element onto the corresponding element of the structural representation. This concludes the definition of the structural representation, which is then persistently stored in the system and can be used as described above. The structural and graphical representations are automatically synchronized after having been associated with one other. Therefore, it is not relevant whether the user selects an element of the structural graph or the corresponding graphical element. Either way, the same structural context is set.

The structural representation can be refined or extended at any time, for instance to allow for a finer structural representation or for an additional or a more detailed graphical visualization.

In most cases the user will want to attach additional descriptive semantical properties, i.e. metadata, to the object(s) represented as well as the stored information and documents. Metadata are always structured as a hierarchy, with terms located near the top being more general than terms located near the bottom. Internally, metadata are organized as an ontology, which forms the second component of the MonArch information model (Fig. 9).

There are several ways to add semantical properties to a MonArch database. One can, for instance, upload a normative set of metadata such as the Integrated Authority File from an external source.[7] The system allows the user to define his own proprietary set of metadata. This was the primary source of metadata in early MonArch projects (see the St. Lorenz examples). It is also possible to refine a stored set of metadata by adding properties that are more specific. However, this third option has to be used with care.

Once the first two components of the MonArch information model, i.e. the structural representation and the set of semantical properties, have been established, target information can then be inserted into the MonArch database. Basic information, for instance length, width and height or geo-coordinates, can be attached to the object or its structural parts. Documents and other digital objects can be assigned to arbitrary elements in the structural representation of the building or site. Additional information and documents can be inserted at any time as needed. Multiple assignments are possible.

There exist two ways of assigning information and documents to structural elements: by following a structured insertion dialogue, i.e. a wizard, or by simple drag-and-drop. Bulk loading of an entire collection of documents is also possible. Assigning metadata to documents or structural elements works analogously, i.e. either by following a structured insertion dialogue or by drag-and-drop.

The insertion wizard leads the user through the following steps of an insertion dialogue: assignment of formal metadata such as time of creation, author etc.; assignment of descriptive semantical properties according to the vocabularies available; and assignment of the data object to be inserted to suitable elements of the structural representation. Finally, a preview of the data object inserted is provided. The entire dialogue or parts of it can be reiterated any time, e.g. for correction or for specifying more detailed properties.

Almost any kind of digital information in any common data format can be stored in a MonArch database. Examples include digital documents such as digitized historical maps and plans, CAD drawings, 3D models, photographs, text documents, climate data, stream based data, spread sheets, sound files, and videos. Even links or signatures relating to a physical document or item can be stored.

After having inserted information and documents, the MonArch database can be searched by selecting a structural element. Information about the structural element selected can be obtained and all documents conveying information about this element can be viewed. The user can refine the search by specifying semantical properties forming an additional search filter. The documents found can be viewed or processed using standard editors, players, viewers etc., as provided by the computer system.

In addition, the MonArch system provides a report generator, which is flexibly adaptable to specific requirements, e.g. the generation of room data sheets. Additionally, the information and documents found can be exported for further use. Document versions are supported, images can be exported at different scales, the graphical representation can be exported as CAD-graphics, and metadata can be exported.

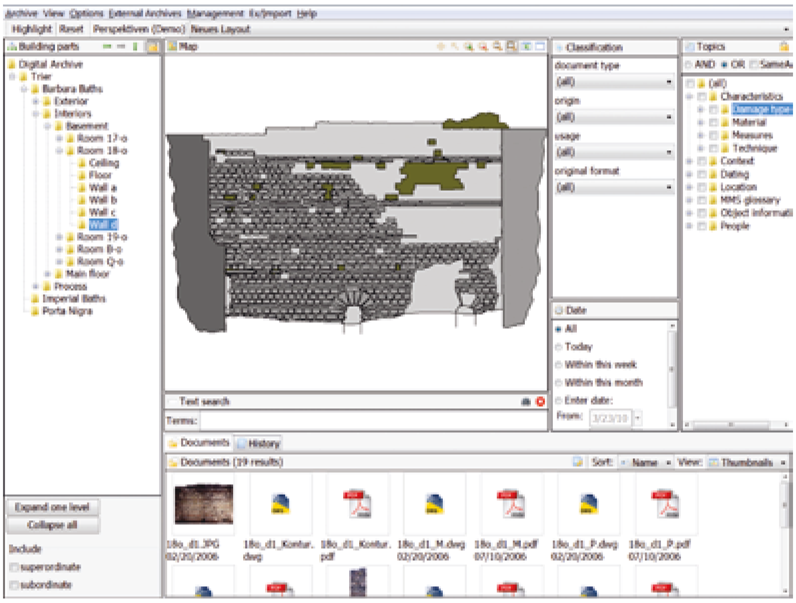

Fig. 8: Detailed map displaying damaged parts of wall

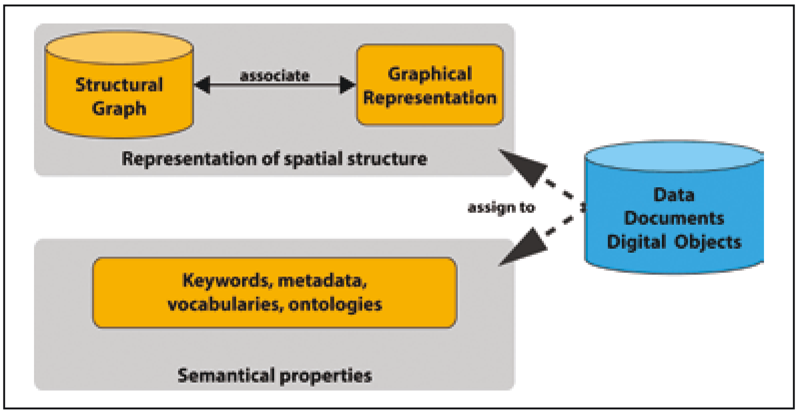

Fig. 9: MonArch information model consisting of three major components

There are at least two major advantages to the MonArch approach. The first and rather obvious advantage results from the principle of organization by structure, which provides a homogeneous form of accessing the stored information. Relevant documents can easily be systematized even if they are widely scattered over many locations. Even infor-

Properties of the MonArch approach

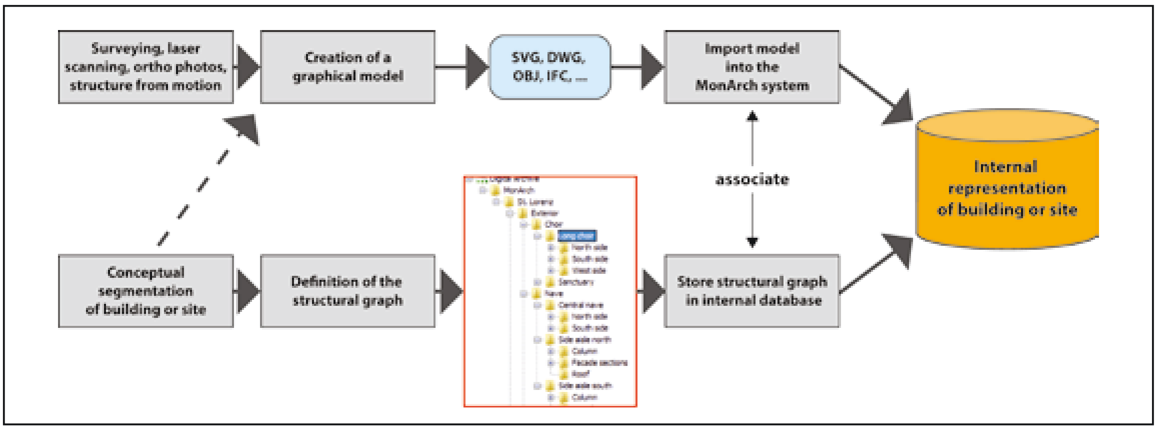

Fig. 10: MonArch set-up workflow

mation that does not seem to be related to a (sub-) structure, e.g. finances, work processes or scientific findings, can be arranged in a suitable way. A major reason for its ease of use is the fact that all information stored is given a structural and semantical context, which resembles the perspective of many users working in the fields of preservation, maintenance, art history, the history of architecture and construction, and other areas. In summary, MonArch ensures intuitive, straightforward access to the relevant information. Once a digital model of the building or site under consideration has been created, using a MonArch database does not require deep technical training or skills in computer science. MonArch therefore allows users to remain in their normal realms of work when working with the digital representation of a building or site.

There is also a second, less obvious but nevertheless important advantage. When using MonArch there is a central point of access to all relevant information and documents. Different applications can share information and documents. Metadata and structural representation can be shared between different applications. Selected contents of a MonArch database can be shared with the general public via the Internet. Moreover, MonArch provides consistent and complete information for different applications and use cases. Different aspect angles, e.g. of preservation, archaeology, building maintenance, restoration, and residential use, can be supported by one body of information. It is not necessary to duplicate the information for each and every application. Thus information consistency and completeness (in a sense) is maintained. In summary, MonArch supports shared and collaborative work based on a common body of information. As described below, collaboration is even possible with remote partners over a network.

Technical aspects

The MonArch system is based on current database technology. Therefore, even huge data volumes can be stored and managed. As a true multi-user database system, it provides access control, data protection and simultaneous usage of the information base by many users.

Access to documents and metadata in MonArch is controlled by a role-based access control system. The structural model, the target information and metadata can be selectively protected. An existing user and group management can be integrated.

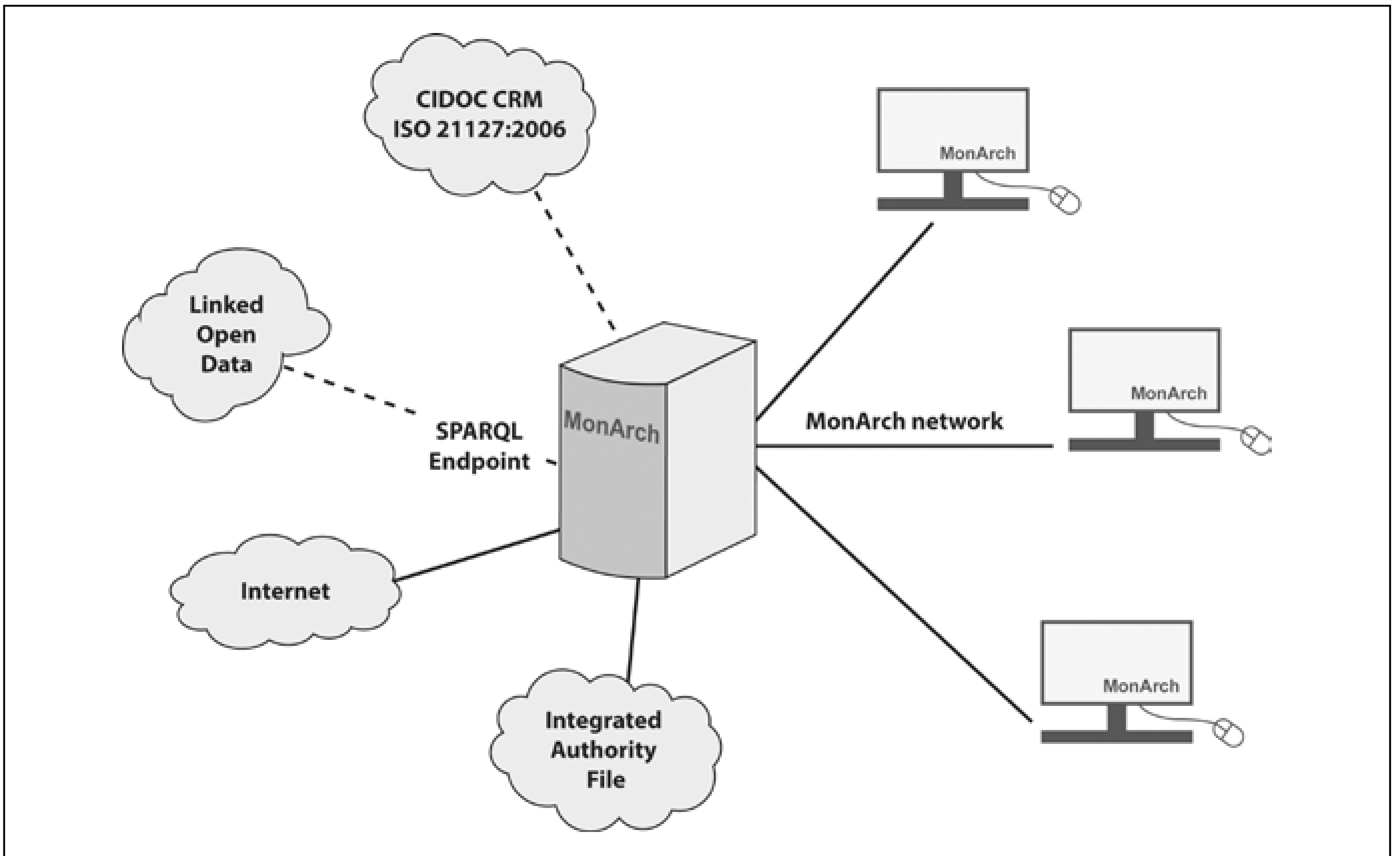

Fig. 11: MonArch network

MonArch has a client-server architecture and can be part of a network. Multiple MonArch clients can connect to a server that manages and stores the representation of the object(s) and the related information. The networked software architecture allows for optional professional hosting of the core system and data while still providing maximum flexibility to the user. Moreover, multiple MonArch servers can be connected via a network, thus allowing collaborative work and queries spanning more than one object, e.g. when working on closely related cultural objects (Fig. 11).

The following interfaces (among others) are supported: DXF for CAD drawings used for graphical visualization, GML for the exchange of spatial objects, Blender for 3D models, RDF for general metadata.

The MonArch software is the outcome of a series of academic projects. It is free to use in non-profit research projects. Software maintenance and development continues, and user training and assistance in projects both academic and professional are available.

Recapitulation and outlook

MonArch supports sustainable storage and management of all information and documents related to a building or site. Using a MonArch database does not require deep technical training or skills in computer science.

One direction of current research and development is devoted to finding a broader and more flexible way of integrating a MonArch database into the global information network. To this end, the capability of integrating external vocabularies will be extended and ways to link MonArch information items to external information sources in the form of linked data[8] will be incorporated.

Concrete development work is underway towards implementing an interface to CIDOC CRM (ISO 21127:2006)[9] as a standard for the description of cultural objects; also under development are a SPARQL endpoint[10] and an interface for 3D models according to the IFC standard (ISO 16739:2013).[11]

Illustrations

1–11 MonArch

Annotations

1 www.monarch.uni-passau.de/en/ (15 February 2017).

2 Drewello, Rainer / Freitag, Burkhard / Schlieder, Christoph: New Tools for Ancient Ruins, in: DFG Magazine German Research, vol. 33, issue 1, 2011, pp. 24–28.

3 Stenzer, Alexander / Woller, Claudia / Freitag, Burkhard: Monarch: Digital Archives for Cultural Heritage, in: iiWAS 2011: Proceedings of the 13th International Conference on Information Integration and Web-based Applications & Services, Ho Chi Minh City 2011, pp. 144–151.

4 www.uni-bamberg.de/en/buchara-projekt/ (15 February 2017).

5 Dodt, Michael / La Torre, Martino: Neue Untersuchungen und Restaurierungen an den Trierer Kaiserthermen, in: Bericht über die 48. Tagung für Ausgrabungswissenschaft und Bauforschung, Erfurt 2014, pp. 197–206.

6 arch.rwth-aachen.de/cms/Architektur/Forschung/Verbundforschung/Cultural-Heritage/~cbjv/Pfalzenforschung-in-Aachen/?lidx=1 (15 February 2017).

7 en.wikipedia.org/wiki/Integrated_Authority_File (15 February 2017).

8 www.w3.org/standards/semanticweb/data (15 February 2017).

9 www.cidoc-crm.org (15 February 2017).

10 www.w3.org/TR/rdf-sparql-query/ (15 February 2017).

11 www.buildingsmart-tech.org/specifications/ifc-overview (15 February 2017).