

## CONCEPTS AND TECHNOLOGIES FOR A COMPREHENSIVE INFORMATION SYSTEM FOR HISTORICAL RESEARCH AND HERITAGE DOCUMENTATION

F. Henze<sup>a,\*</sup>, N. Magdalinski<sup>b</sup>, F. Schwarzbach<sup>b</sup>, A. Schulze<sup>c</sup>, Ph. Gerth<sup>d</sup>, F. Schäfer<sup>d</sup>

<sup>a</sup> BTU Cottbus, Chair of Surveying, 03046 Cottbus, Germany – [henze@tu-cottbus.de](mailto:henze@tu-cottbus.de)

<sup>b</sup> HTW Dresden, Faculty of Spatial Information, 01069 Dresden, Germany –  
([magdalinski,schwarzbach@htw-dresden.de](mailto:magdalinski,schwarzbach@htw-dresden.de))

<sup>c</sup> BTU Cottbus, Chair of Database and Information Systems, 03046 Cottbus, Germany – [schulal2@tu-cottbus.de](mailto:schulal2@tu-cottbus.de)

<sup>d</sup> German Archaeological Institute, 14195 Berlin, Germany – [philipp.gerth@dainst.de](mailto:philipp.gerth@dainst.de), [felix.f.schaefer@gmail.com](mailto:felix.f.schaefer@gmail.com)

**KEY WORDS:** Information Systems, Digital Documentation and Publication, Data Retrieval, 3D GIS, Archaeology

### ABSTRACT:

Information systems play an important role in historical research as well as in heritage documentation. As part of a joint research project of the German Archaeological Institute, the Brandenburg University of Technology Cottbus and the Dresden University of Applied Sciences a web-based documentation system is currently being developed, which can easily be adapted to the needs of different projects with individual scientific concepts, methods and questions. Based on open source and standardized technologies it will focus on open and well-documented interfaces to ease the dissemination and re-use of its content via web-services and to communicate with desktop applications for further evaluation and analysis.

Core of the system is a generic data model that represents a wide range of topics and methods of archaeological work. By the provision of a concerted amount of initial themes and attributes a cross project analysis of research data will be possible. The development of enhanced search and retrieval functionalities will simplify the processing and handling of large heterogeneous data sets. To achieve a high degree of interoperability with existing external data, systems and applications, standardized interfaces will be integrated. The analysis of spatial data shall be possible through the integration of web-based GIS functions. As an extension to this, customized functions for storage, processing and provision of 3D geo data are being developed.

As part of the contribution system requirements and concepts will be presented and discussed. A particular focus will be on introducing the generic data model and the derived database schema. The research work on enhanced search and retrieval capabilities will be illustrated by prototypical developments, as well as concepts and first implementations for an integrated 2D/3D Web-GIS.

### 1. CONTEXT AND MOTIVATION

In the context of current archaeological research projects, involving various historical and neighbouring scientific disciplines, usually large amounts of digital data are produced or collated. However, in antiquity studies adequate solutions for a consistent and project independent documentation, delivery and archiving of digital research data are still missing. Usually domain, project or institution specific systems are created and used independently. Often these solutions cannot be transferred to other projects, despite overlapping content. This leads to a regular high effort for providing project-specific tools and the large number of different solutions complicates the exchange and long-term availability of digital research data.

In addition, the generated data for one object are often held in different systems and data formats, depending on the disciplines involved. Information is generated multiple times independently of each other and is stored in different versions and working drafts. This inevitably leads to inconsistent data within a project and complicates a cross-project analysis. Due to lack of concepts for long-term provision of research data beyond the end of a project, only a small portion of the digital data is usually published. The greater part is not available for further research and finally falls into "digital oblivion".

To avoid this heterogeneity two information systems have been developed in the context of different research projects at the

German Archaeological Institute (DAI) and at the Brandenburg University of Technology Cottbus (BTU) during the last years:

- CISAR - a web-based information system for archaeology and building archaeology at the chair of surveying at BTU Cottbus, <http://www.tu-cottbus.de/cisar/>, (Henze et al, 2009)
- iDAI.field - a documentation system created at the IT department of the DAI, <http://www.dainst.org/de/project/idaifield>, (Schäfer, 2011).

Although CISAR and iDAI.field have largely proven in numerous archaeological projects over the last 6 years, they have some fundamental deficits. Thus, tools and methods for a cross-project exchange of research data are missing. Networking with online resources or with external databases is not possible due to lack of interfaces. Because of incomplete and inflexible data models and outdated software architectures adapting to new research projects as well as the general maintenance and the further development of the systems require a very high effort. Both systems lack a fundamental concept that describes their long-term use, development and funding beyond the end of research projects. In addition, opportunities to secure a long-term digital publication of already released research data are missing. Thus, the two solutions above are not suitable to ensure the current technical and domain-specific requirements for an adequate provision of research data.

---

\* Corresponding author.



### 3.2 Data Types for Attribute Values

Because of the generic design of the database schema preferably all attribute values are stored in the same relation to the SQL data type "text" regardless of the associated *Attribute Type*. This approach allows for dynamic storage of attribute values that normally require different SQL data types, without making any changes to the database schema. Thus, however, further adjustments are required to maintain the original SQL data type of the attribute value. The SQL data type itself must be saved.

This is done within the relation *Attribute Type* on the column *Data Type*, which is stored as alphanumeric value as part of a *Value List* (see chapter 3.4). If a SQL data type specific function is used a cast to the appropriate *Data Type* must be made in advance. This cast can be realized only through a SQL query generated in the program logic or through a prepared database view.

However, an exception must be made for the storage of the PostGIS data type "geometry". Attached Geographic Information Systems (GIS) query large amounts of geometry data for presentation and analyses. Through additional casts on "geometry" data type, major delays are expected. Because of this performance issue, we transferred all "geometry" data types to a separate relation.

### 3.3 Multilingualism

For implementation of multilingualism, a concept was used, which is known in geoinformatics under the term *PT\_FreeText* (see ISO 19115-1). The corresponding UML modeling is shown in Figure 3. All languages available in the system are described in the class *Language Code* by a code according to ISO 639-2. In the class *Character Code* corresponding character codes for texts according to ISO 19115 are provided and the class *Country Code* includes all country descriptions according to ISO 3166. There are default entries with the values "deu" (language code), "DE" (country code) and "UTF-8" (character encoding). Finally, the class *PT\_Local* brings together the equivalent encodings. The translation entries themselves are stored in the class *PT\_LocalizedCharacterString*, which in turn connects to the class *PT\_FreeText* and thus ensures the connection to the actual classes that contain the data.

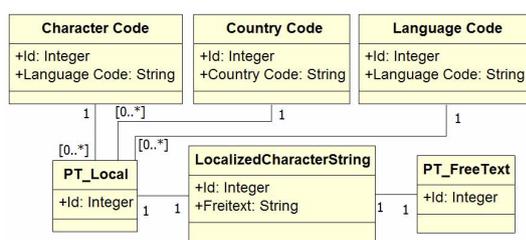


Figure 3: Multilingualism in UML notation

For the realization of multilingualism the columns *Name* and *Description* were identified in the database design. If present, these columns were replaced in respective relations by foreign keys to the relation *PT\_FreeText* (see Figure 1 and 2).

### 3.4 Value Lists

*Value Lists* (VL) are an important part of the dynamics of OpenInfRA. They extend code lists (ISO 19103) used in geoinformatics by enabling hierarchy and visibility. Each *Value List* can contain *n* values. Via *visibility*, these values can be shown or hidden project-specifically. Further they can be set in

relation to each other using SKOS\* relationships, thus ensuring a high flexibility in the design of individual projects. *Value Lists* are managed in the system database and subsequently distributed across projects. Thereby, redundancy in *Value Lists* is avoided but availability in the projects may be delayed. However, over time, a comprehensive set of *Value Lists* and corresponding values will have been developed and further enhancements will rarely be needed.

## 4. SEARCH AND RETRIEVAL STRATEGIES

Information systems that are based not only on databases, but also on files and documents are extremely useful for saving huge amounts of data. But without corresponding techniques to retrieve this data, the stored information will not be sufficiently accessible. The intention is to combine the power of a relational database and the query language SQL with the impreciseness and proximity of the information retrieval. To achieve this, we will draw on a framework for uncertain retrieval on certain data that is called *Commuting Quantum Query Language (CQQL)* and is implemented as *Quantum SQL (QSQL2\*\*)*. In addition to many possible retrieval methods and filters, we will also try to keep the Graphical User Interface (GUI) as smart as possible.

An archaeological question that exemplifies the complexity of the task would be as follows:

"Find all sculptures which have the original height (preserved or reconstructed) of around 1.80 m (e.g. around life-size) and search not only in the actual database for stone finds but look also in the old excavations reports which have been attached as PDFs to the excavations datasets."

### 4.1 Conceptual Design of the Retrieval System

In order to address a wide range of users, we plan to integrate different types of retrieval functionalities. Technically, the resulting retrieval systems will differ by the way they look and how they search the data. We can divide them into *text* and *SQL based retrieval systems*.

The *text based retrieval system* will hold an inverted index (Heinrich, 2008) on the actual data that will be used to answer a user request. The database and the documents will be indexed by the system in a defined interval. The system will support different analyzers like stop word filter, automatic language detection, synonyms (based on SKOS), stemmers and the support of different search operators and placeholders. To give the user access to this index we provide two kinds of interfaces that can be used to formulate a request. The first interface is a simple input field called *basic search* that can be found on every page of OpenInfRA. For more detailed research the basic search can be extended to an *advanced search*. It offers more visual options for defining a search query or restricting the search domain. Every extension of the request that will be formed in the advanced search can also be formulated in the basic search by using special keywords. This retrieval system will appeal to users that are not familiar with the structure and the content of OpenInfRA.

The *SQL based retrieval system* will use the QSQL2 language. This retrieval system is addressing expert users being familiar with the OpenInfRA data structure. To generate a SQL based

\* SKOS (Simple Knowledge Organization System), visit <http://www.w3.org/2004/02/skos/> for further information.

\*\* QSQL2 is an extension of the logic-based query language QSQL (and therefore of SQL). Further information about QSQL can be found in (Lehrack and Schmitt 2010).

request no knowledge about SQL or QSQL2 is needed. We will provide two interfaces. The first one is called *detailed search* and was already used in the legacy system iDAI.field. This interface uses the same form for query building and for data entry. Furthermore, options such as calling index lists for quick access to possible input or weighting parts of the request will be available. The second one will be a more database-oriented solution and is called *expert search* working like a construction kit for database queries. The user can take several visualized parts of the database and put them together to build a statement. Furthermore, the generated QSQL2 / SQL statement will be directly modifiable.

For every request in the SQL based system, the user can choose to make use of the uncertain functionality provided by QSQL2. If only certain data shall be queried, this will be done by standard SQL and thus produce no additional performance overhead.

#### 4.2 Result Views

The user will be able to choose between four different query result views. Each view is initially sorted by the relevance of the results that were delivered by the retrieval system and that are displayed as bar charts.

(i) *list-view*: The standard view for all results is a list view that can be compared with the result list delivered by Google. The results cannot be sorted by the user.

(ii) *catalogue-view*: This view shows more details of the results but is still compact. Each project can predefine a specific selection, order and formatting of a fixed number of attributes that will be shown for each dataset (e.g. a picture, an object-id or a document name, a title or a long description). This selection also determines the order of the results.

(iii) *table-view*: A table-like presentation with attributes arranged in columns and datasets in rows. Users can change the range of displayed attributes for each result. It is possible to add and hide the shown attributes, limited to a minimum of two attributes. It will be possible to sort the results list by one of the attributes.

(iv) *thumbnail-view*: This view is specialised for results that contain pictures, otherwise the results will be filtered out. Instead of showing text from attributes in a list, the results will be shown as tiles. Only one or two textual attributes will be associated with a thumbnail (e.g. short description and photo number). By hovering over or clicking on a picture, it will be zoomed and more information will appear.

#### 4.3 Filtering and Export

Additionally, a filtering mechanism allows the setting of several filters already during the construction of a search request. This further reduces the retrieval time. It will also be possible to limit the results in the different result views as described above. To make use of the retrieved data from OpenInfRA, the results can be exported in different formats (e.g. CSV, XLS). The user can choose the data from the result list and furthermore select the attributes of the chosen data sets that will be exported.

### 5. 3D WEBGIS

In addition to collecting, analyzing and providing semantic data in domain specific information systems spatial information are increasingly included into the analysis of data in historical research. Therefore special data types and functions for web-based storage, processing and visualization of spatial data will be made available in OpenInfRA. While web-based GIS

functions for 2D geometries can be implemented largely with standardized services and formats, appropriate techniques for processing and visualization of 3D spatial data have not yet been established. Based on recent developments for web-based 3D visualization and database management of 3D spatial data, 3D GIS functionalities will be developed and integrated into OpenInfRA. The development work is done in close cooperation with the design and implementation of classical 2D WebGIS functions.

#### 5.1 Application Areas for 3D

The scope of use for 3D geometries ranges from the recording and visualization of *individual find objects* on the modeling and analysis of *complex building structures* to the use of three-dimensional *city and landscape models* for the analysis and representation of larger topographical contexts.

(i) By 3D acquisition of *individual objects*, such as small finds, architectural ornaments and fragments, a new quality of documentation can be achieved. Usually a 3D visualization of the objects is sufficient, further GIS analyzes are not required.

(ii) In contrast, 3D models of complex buildings or entire settlements are used for analysis and presentation of spatial contexts and processes, thus going far beyond the requirements of a (simple) 3D visualization. Architectural reconstructions and development phases will be investigated and visualized using detailed 3D models. They are divided into different topographical and/or structural and constructive object parts which are associated with semantic information and with information about the graphical expression and metadata.

(iii) For geometric object recording scanning techniques are used increasingly. The results are very dense 3D point clouds that correspond to an un-interpreted, "neutral" 3D image of the object. This point clouds can be used as an important basis for (future) archaeological investigations and building research. They are therefore also provided in the system and visualized along with other measurement data.

#### 5.2 Data Management and Data Modeling

Similar to 2D or 2.5D geometries, 3D geometries are stored in a spatial database, which must offer special 3D data types and functions. With the new 3D support of PostGIS 2.0 it can be used along with PostgreSQL as geo-database for OpenInfRA. It has yet to be investigated if the existing PostGIS functions for validation, storage and processing of 3D geometries meet the requirements of archeology and building research or whether they must be supplemented by application-specific functions.

For modeling of 3D geo data, CityGML for management of 3D city models (<http://www.citygml.org/>) and IFC for building information modeling (<http://www.buildingsmart.org/standards/ifc>) have been established as standardized solutions. Both approaches use a common modeling of geometric-topological and semantic data (spatially structured and semantically coherent objects). They are therefore contrary to the generic data model for OpenInfRA.

A pragmatic approach for geometric data modeling has been developed for the Building Information System (BIS) for Domus Severiana in Rome (Brasse et al, 2009). The building structure is first modeled only on semantic data by creating "data sheets" for all structural and constructive elements. Parallel to this, a volume model is created, whose geometric elements are assigned as additional text attribute to the semantic data in form of VRML code. The approach offers a high degree of flexibility in the assignment and modeling of 3D geometry and it is not fixed to a specific geometry model (volume vs.

surface model). However, since the geometry is stored only as VRML code, spatial queries are not possible via the database. Using current technologies, like 3D Geo-Database, X3D and WebGL (see below), the data model can be applied to OpenInfRA and converted into a standardized 3D WebGIS application.

### 5.3 Prototypical Implementation

Essential requirements for the application are the standard compliance of all services and formats (OGC and W3C standards) and the use of open and fully documented software (open source). The application should run without additional software installation on commonly used Internet browsers and on different operating systems. While 2D WebGIS are realized largely using OGC standards, relevant specifications for the 3D data are still in development. Similar to standard 2D services (e.g. WMS, WFS, WCS, <http://www.opengeospatial.org/standards>), an OGC-Draft for a Web 3D Service (W3DS, version 0.4; OpenGIS, 2010) exists, that is used in OpenInfRA as basis for the 3D WebGIS. As exchange formats the service uses X3D and KML/Collada i.a. X3D is an XML-based graphics format that was adopted by the Web3D Consortium as official standard for 3D content on the Internet (<http://www.web3d.org/x3d/specifications/>). With the "Geospatial Component", a concept exists for X3D to support geographical and geo-spatial applications. It supports reference coordinate systems, levels of detail (LOD) and geographic coordinates support i.a.

#### First Realization of a 3D Client

In a first step a 3D WebGIS client has been developed that meets the requirements, but not yet communicates over an OGC-compliant service (see Figure 4, phase 1). The client application was implemented using HTML5 and CSS3. Functionality for interaction (layer selection, navigation modes, model color, etc.) have been implemented in JavaScript. The 3D models have been converted to X3D and uploaded to the Web Server as static X3D files. The conversion of X3D models into the declarative scene description used in HTML5 for hardware accelerated visualization with WebGL (<http://www.khronos.org/webgl/>) is carried out using the JavaScript framework X3DOM (Behr et al, 2009, see also <http://www.x3dom.org/>).

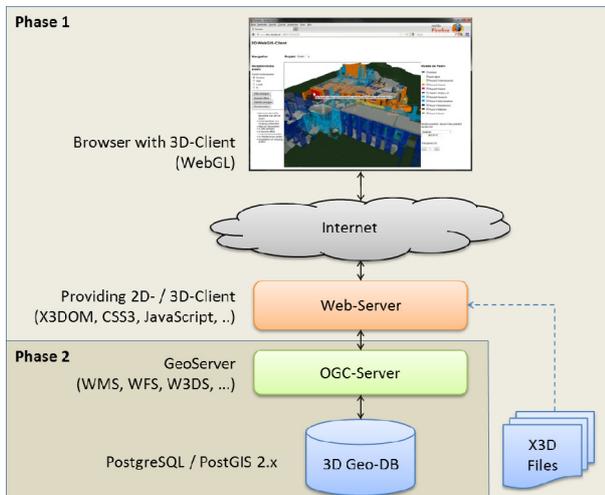


Figure 4: Client/Server architecture for the realization of a 3D WebGIS

By the use of these components, the client application is platform independent and can be implemented entirely using freely available software. The user just needs an internet connection and a recent browser installation. Via this browser the client URL is called and as response the web server returns the application. Both, the interactions and the representation of 3D geometries performed locally on the user's PC. The requested X3D file is completely downloaded from the server before the geometry is rendered in the browser via WebGL. The Web Server will always be contacted when a new 3D model has to be added. Figure 5 shows the user interface of the prototype of the first realization phase.

By using the X3DOM framework, basic functions for viewing and navigation in the 3D models are already available without any further programming work:

- seven different navigation modes for control and orientation of the scene with the mouse.
- a console for documentation of relevant steps of the framework and for error messages
- superimpose of additional information about the frames per second (fps) and the number of rendered points and polygons
- three rendering modes for the display of points, wireframes or shaded/textured polygons

In addition to these basic functionalities own extensions were developed:

- Interactive adding of models or layers: parts of a 3D model or additional models can be added via a checkbox selection or removed from the scene (see Figure 5 on the right). This corresponds to layer selection in 2D web GIS applications
- Change of displayed color and transparency for layers
- Integration of textures and georeferenced images: in addition to generic textures (brick wall, wood, etc.) georeferenced raster data can be integrated for texturing of terrain models with aerial photographs or thematic raster data and buildings or single objects with photogrammetric image textures.
- Query of semantic object information: using a mouse over function, the object under the mouse cursor is highlighted and a tool tip will be shown. Currently, a data-sheet is opened via a static link. In the future, a standardized "GetFeatureInfo" request will be made to the geo-database.

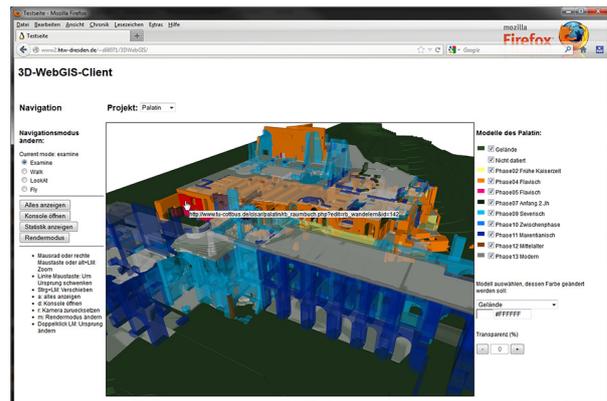


Figure 5: Prototype of the 3D client with the complete building model of the Domus Severiana on Palatine in Rome.

#### Implementation of a Web 3D Service

The above described client prototype functions already cover most of the essential requirements for web-based visualization of 3D geo data. The geometry data, however, are still only static X3D data and thus do not allow typical GIS operations and queries. Therefore, in a second stage the static linked X3D

models were replaced by dynamically generated models from a spatial database (Figure 4, phase 2). As server-side software a customized implementation of the "GeoServer" (<http://geoserver.org>) is used. In addition to 2D mapping services (WMS, WFS, WCS) it contains an implementation of the Web 3D Service (W3DS) according to the current OGC draft specification 0.4 (OpenGIS, 2010). After a "GetScene" request the service accesses 3D geometries in a PostGIS 2.0 database and returns them as X3D data. These data can be visualized in the 3D window of the client application. Initially, simple 3D geometries of buildings from the Baalbek research project (Henze et al, 2009) were used as test data. This data could be integrated without additional programming or customization of the client.

Currently, the functions listed in the W3DS specification draft 0.4 for service-based integration and retrieval of 3D geometries are examined in more detail, and will be integrated into the client application. The result is a standards-compliant 3D WebGIS architecture, which is composed of phases 1 and 2 shown in Figure 4 and which completely corresponds to the architecture examined in the OGC 3D Portrayal Interoperability Experiment (OGC, 2012) consisting of PostgreSQL/PostGIS 2.0, W3DS, X3D and X3DOM/WebGL.

## 6. SUMMARY AND OUTLOOK

The developed database provides a flexible and generic approach in handling data. It supplies standardized multilingualism and the capability to support different characteristics of projects within the same database schema. The development of the database schema is mainly based on theoretical considerations and has been tested with computer generated data only. The next step is the migration of project data from legacy systems into the new database schema. At this point we expect smaller problems, when the theoretical design meets the practical data.

We identified several levels of retrieval in OpenInfRA. To provide different users with adapted retrieval mechanics, we developed four different search masks that make use of text- and SQL-based retrieval. After the migration of real data, the developed data retrieval systems can be prototypically implemented, tested and assessed.

As part of the design phase of OpenInfRA visualization options for 3D geometries in the browser were identified. Based on standard compliant technologies and formats a prototypical web-based GIS architecture, consisting of a 3D geo-database (PostGIS 2.0), a map or geodata service (GeoServer with W3DS) and a 3D client application (HTML5, X3DOM/WebGL) was realized. Previous development work is focused on the implementation of the 3D Web GIS client for rendering X3D data available both, as static files and as standard output of a Web 3D Service. As part of ongoing work, all of the operations implemented in the GeoServer W3DS so far (e.g. GetCapabilities, GetScene, GetTile and GetFeatureInfo) will be examined and made available in the client application. In this context, it should also be investigated to what extent the 3D client can be integrated into existing 2D clients and whether such integration is actually useful. For the integration of point clouds appropriate frameworks (e.g. XB PointStream, [http://zenit.senecac.on.ca/wiki/index.php/XB\\_PointStream](http://zenit.senecac.on.ca/wiki/index.php/XB_PointStream)) will be investigated and integrated into the application. For a web-based data input, interfaces will be provided that enable users to import 3D data into the database. It will be examined which 3D formats are well accepted among users (such as DXF, VRML,

3DM) and how these formats fulfill the requirements of a 3D WebGIS. In addition, functions will be provided to validate imported models in order to avoid errors in processing, analyzing and displaying data.

## 7. ACKNOWLEDGEMENT

The research project OpenInfRA is funded by the "Deutsche Forschungsgemeinschaft" (DFG) since April 2011 within the program "Information Management".

## REFERENCES

### References from Journals:

Henze, F., Lehmann, H., Brusckhe, B.: Nutzung historischer Pläne und Bilder für die Stadt-forschungen in Baalbek / Libanon. In: Photogrammetrie – Fernerkundung – Geoinformation (PFG), Heft 3/2009. S. 221-234.

Lehrack, S. and Schmitt, I., 2010: „QSXL: Incorporating Logic-Based Retrieval Conditions into SQL“. *Computer Science* 5981/2010, pp. 429–443.

### References from Other Literature:

Behr, J., Eschler, P., Jung, Y., Zöllner, M., 2009: X3DOM: A DOM-based HTML5/X3D integration model. In: Stephen N. Spencer (Ed.): *Proceedings of the 14<sup>th</sup> International Conference Web3D '09*, ACM, New York, USA, pp. 127-135.

Brasse, Ch., Heine, K., Zhao, D., Wulf, U., 2009: A 3D Solution for a Web-based Building Information System. In: Posluchny, A., Lambers, K., Herzog, I. (Eds.): *Layers of Perception*, Proceedings of the 35<sup>th</sup> International Conference on Computer Application and Quantitative Methods in Archaeology (CAA), Dr. Rudolf Habelt GmbH Verlag, Bonn.

Heinrich, A., 2008: Information Retrieval 1 - Grundlagen, Modelle und Anwendungen, Otto-Friedrich-Universität Bamberg.

Henze, F., Lehmann, H., Langer, W., 2009: CISAR - A Modular Database System as a Basis for Analysis and Documentation of Spatial Information. In: Posluschny, A., Lambers, K., Herzog, I. (Eds.): *Layers of Perception*, Proceedings of the 35<sup>th</sup> International CAA Conference, Dr. Rudolf Habelt GmbH Verlag, Bonn, pp. 228-233.

Schäfer, F. F., 2011: iDAI.field - Ein modulares Dokumentationssystem für archäologische Feldforschungsprojekte. *Kölner und Bonner Archäologica (KuBA)* 1, Köln/Bonn, pp. 193–200.

Schulze, A., Schäfer, F., Gerth, Ph., Schwarzbach, F., Henze, F., 2012: OpenInfRA – Storing and retrieving information in a heterogeneous documentation system. In: *Proceedings of the International CAA Conference*, University of Southampton, March 2012 (in press).

### References from Websites:

OpenGIS Discussion Paper, 2010: Draft for Candidate OpenGIS® Web 3D Service Interface Standard (Referenznr. OGC 09-104r1), [http://portal.opengeospatial.org/files/?artifact\\_id=36390](http://portal.opengeospatial.org/files/?artifact_id=36390) (8. Jul. 2013)

Open Geospatial Consortium (OGC), 2012: 3D Portrayal Interoperability Experiment, <http://www.opengeospatial.org/projects/initiatives/3dpi> (8. Jul. 2013)