
The Arachne Object-database

**Ortwin Dally, German Archaeological Institute, and Reinhard Foertsch,
University of Cologne (See [email contacts page](#) for the authors' email addresses.)**

[Arachne](#) is the central object-database of the German Archaeological Institute. In 2004, the German Archaeological Institute (DAI) and the Research Archive for Ancient Sculpture at the University of Cologne (FA) joined their efforts in providing and further developing Arachne as a tool for free internet-based research.

Arachne's database design uses a world-model that tries to build on one of the most basic assumptions one can make about archaeology, classical archaeology or art history: all activities in these areas can generally be described as contextualizing objects. When researching potential project models, we found that some had more complicated but less compatible world models. The model chosen by Arachne allows for a general information-retrieval of a powerful pool of material, whereas on the level of category-properties, very specific structures can be displayed.

Thanks to significant and ongoing support by the Deutsche Forschungsgemeinschaft, Arachne started to integrate negative archives of ancient sculpture from sources with large collections such as the German Archaeological Institute in Rome and the historic glass-negative collections of the German Archaeological Institutes in Athens, Cairo, and Istanbul.

As of February 2007, Arachne contained about 1703 registered users who can access 161,000 scans and 112,000 objects free of charge. Twenty to thirty years from now, about 700,000 images from the DAI could be expected in Arachne. With the addition of newly produced documentation, one could hope for 1 million objects.

Interoperability is key to how data are shared and connected, mainly between Arachne and several GIS systems used on DAI excavations and surveys. A given object will only reside once inside the overall dataspace of DAI. Uniform resource names will unmistakably identify objects residing in Arachne. DAI and FA plan to implement Syntactic Interoperability for its data resources, for which a collaboration with the Perseus Digital Library and its Editor in Chief, Greg Crane, is ongoing. The project scope as of February 2007 can be followed on a personal page of Robert Kummer inside the Perseus Wiki (<http://devwiki.perseus.tufts.edu/wiki/User:Robert>).

Due to a lack of effective access to existing data sources that are spread all over the world some scientific questions cannot be addressed today. The main goal therefore is to identify, describe and implement elements of a cyberinfrastructure for classics and archaeology. This cyberinfrastructure would provide unified web-based and multilingual access to distributed resources that have been integrated syntactically and semantically.

Syntactic interoperability could be provided by the use of the Collection Services Protocol, the OAI (Open Archives Initiative) Protocol for Metadata Harvesting or a simple RSS/Atom-Feed. These protocols are suitable to disseminate effectively any kind of metadata that can be expressed in terms of XML through a standardized interface.

One major aid for semantic integration will be the CIDOC Conceptual Reference Model. (CIDOC is the International Committee for Documentation of the International Council of Museums, ICOMOS.) It will help to analyze the data structures of the participating databases and

to identify common information contents. Furthermore it will provide the standards needed to implement an infrastructure that enables seamless data integration by using a shared metadata format. Once consolidated and indexed in a central repository, this metadata format will support sophisticated indexing as well as multilingual and associative queries.

Since the CIDOC CRM provides a top level framework, certain communities will need to extend the basic vocabulary with their own terms. Therefore - as a second major part - the concept of the CIDOC CRM itself relies heavily on other forms of shared infrastructure. Gazetteers and other domain-specific naming authorities that provide registries of controlled vocabularies will function as semantic registries that provide further information on how the shared data should be processed. These registries will allow users to interpret the retrieved data and application developers to implement software that interprets the data according to a certain goal. The registries have to be developed and published to a great audience to enable reuse and exchange of meaningful data elements. Furthermore semantic registries will tie together all participating databases and play a major role in data discovery.

Traditionally museums present art objects isolated from their contexts and according to specific curatorial decisions. One opportunity that arises by working with the CIDOC CRM is the re-contextualization of those objects by meaningfully connecting them to other art objects of different kinds, to ancient texts for example. This approach permits us to lay the stress on conceptual similarities among objects of classics and archaeology, and it allows the user to discover objects and to navigate from one object to the other based on meaningful links.

One fundamental paradigm to keep in mind is the notion of a process-oriented view of this infrastructure. Leveraging the interoperability capabilities from the metadata to the resource level means supporting scholarly workflows like publication, citation and archival of resources, not just information retrieval. An effective cyberinfrastructure will provide functions for discovery, reference, dissemination, aggregation and other forms of reuse and exchange of resources while preserving intellectual property rights.

The work currently focuses on using the CIDOC CRM for the breakdown of all participating data models by acting partly as a knowledge engineer who interprets the models and the actual data to discover common information content. CIDOC CRM metadata will be compiled for sample data objects and an algorithm will be developed automatically to generate this kind of metadata from currently existing data models. At the same time we are trying to enhance our understanding of the needed architectural elements of a cyberinfrastructure, to identify and to describe them.

On the technical roadmap for the next years are: the integration of specific material categories, which entails the expansion of the Arachne-thesaurus; and improved treatment of search results and search history. Arachne is no exception to general problems of displaying huge search results, and will have to integrate its context-browsing feature with a topic-map design, allowing the user to switch between graphical and more textual displays of information.

If there are other organizations interested in collaborating with the DAI project, please contact: Prof. Dr. Reinhard Foertsch. (See [email contacts page](#) for Prof. Dr. Foertsch's email address.)

-- Ortwin Dally and Reinhard Foertsch

For other *Newsletter* articles [on the use and design of databases](#) or [on the use of electronic media in the humanities](#), consult the Subject index.

[Next Article: Many Heads Are Better Than One](#)

[Table of Contents for the Spring, 2007 issue of the *CSA Newsletter* \(Vol. XX, no. 1\).](#)



[Table of Contents for all *CSA Newsletter* issues on the Web](#)

[CSA Home Page](#)