Cidoc CRM in Data Management and Data Sharing
Data Sharing between Different Databases

Florian ECKKRAMMER[1] – Rainer FELDBACHER[2] – Tobias ECKKRAMMER[2]

[1]FH Technikum Wien
[2]University of Vienna
{florian.eckkrammer; tobias.eckkrammer}@technikum-wien.at
rainer.feldbacher@univie.ac.at

Abstract

In spite of the benefits of globalization research, the integration of the potential of archaeological platforms spanning over multiple projects and excavations is rarely employed due to self-imposed limits. Scientists must be convinced of the advantages of shared approaches to store, describe and maintain data obtained from excavations, findings in particular, publications and interdisciplinary information. The user (data owner) is still in full control of the data exchange process and can decide which data to share with other selected users. Hints should be given as to what kind of finds and contexts have been found. The more members and more different disciplines join, the broader the network and the greater the opportunity for cross disciplinary research. A complete linkage makes direct communication with others or the combing of excavation reports obsolete. The answers to questions queried concerning examinations would be covered by means of metadata.

Keywords

Data exchange, database, data mapping, CIDOC CRM, data accessibility, data availability, archaeological excavations, description lists, peer to peer data exchange, datasharing

1. Introduction

Archaeologists have one crucial problem – they are currently not able to search all the existing information about the topic they are working on at the moment. Conversely, they are not able broadcast their new findings and results of research to other colleagues. As a result they are condemned to search a bunch of new publications for novelties every month. If important information for their studies is published in a range of books they do not search (either they do not know it or cannot access it) they will most likely miss it or will discover it with a great delay.

The ongoing globalization affects all areas of science. This results in cumulative networking and in a broad information and knowledge exchange. The exchange does not only take place in one field of knowledge; it is more like an exchange over several domains. The other domains can be used to provide alternative perspectives or to contribute some technical methods. The results will allow a greater breadth of understanding than a monolithic approach based on the techniques, methods and knowledge of one field. These steps have already been done by archaeologists. This could be seen in (just to name a few exsamples) the employment of 3D laser scanner on excavations, digital storage and recording of blueprints (CAD systems) and the identification via dendrochronology as well as the usage of different databases.

As long as these criteria are not achieved, it will not be possible to avail the benefits of globalization in the field of archaeology.

It is not the intention of the article to tackle all these problems nor would it be possible. All the problems and approaches to solution described in this article refer to the data management and documentation of archaeological excavations. To express the intention of the authors in an abstract way, it could be said that the objective is to create a holistic method for capturing, managing and releasing the data of excavations under consideration of the knowledge of other projects as well as the exchange between them. To get to this point it is necessary to make the data exchangeable and available to others. In succession it would be possible to cumulate the data over a large scale of projects for semantic agents. These agents could search the data pool. To reveal the ancient macrocosm, the self-imposed limits must be forgotten in order to look at new horizons, as even distant regions influence each other due to trade, expansion and resulting conflicts. Therefore it is possible to identify new coherences and so consequences could be drawn in a larger context.

But in the actual situation research is still kept rudimentary and fragmentary. The tools currently used do not provide the functionality. At the beginning of the article the most common techniques for storing data in archaeological excavations will be listed and discussed. Afterwards six critical problems will be isolated and the approaches visualized.

2. Different methods of documentation in excavations

Based on an analysis carried out by F. Eckkrammer *et al.* (2007) of different excavations organized by different countries such as Austria, Germany, Denmark, Sweden, Canada and the UK, the methods of documentation could be categorized as follows:

- Non-computer-based documentation
- Documentation with text processing SW
- Documentation with spreadsheet SW
- Documentation with database SW
- Server centered database

In this probe the criterion for the groups was the system used for saving the data. The documentation was done based on the following areas:

- Textual Documentation
- Drawings & measuring data
- Photos[1]
- Integration of ancillary sciences

This results in the following six core problems related to the attributes of exchangeability and ease of access:

- Different database systems
- Different data structure
- Different description lists
- Different language
- Different accessibility
- Lack of knowledge about relevant DB's

3. Description of main problems

3.1. Different database systems and data structure

Currently different database (DB) systems are used. The number of database platforms used is equal to the number of vendors of such systems. The data of many projects are stored in word processing or spreadsheet software. But processing data stored in such systems is – similar to handwritten information – very challenging and additional transformations are necessary in order to integrate them into the later described workflow.

The different data structure results not from the choice of the DB system or other technical conditions but from the stored findings and data. If for example a database predominantly contains pottery, the lists for categorization of the pottery would be worked out in a very detailed way. On the other hand the

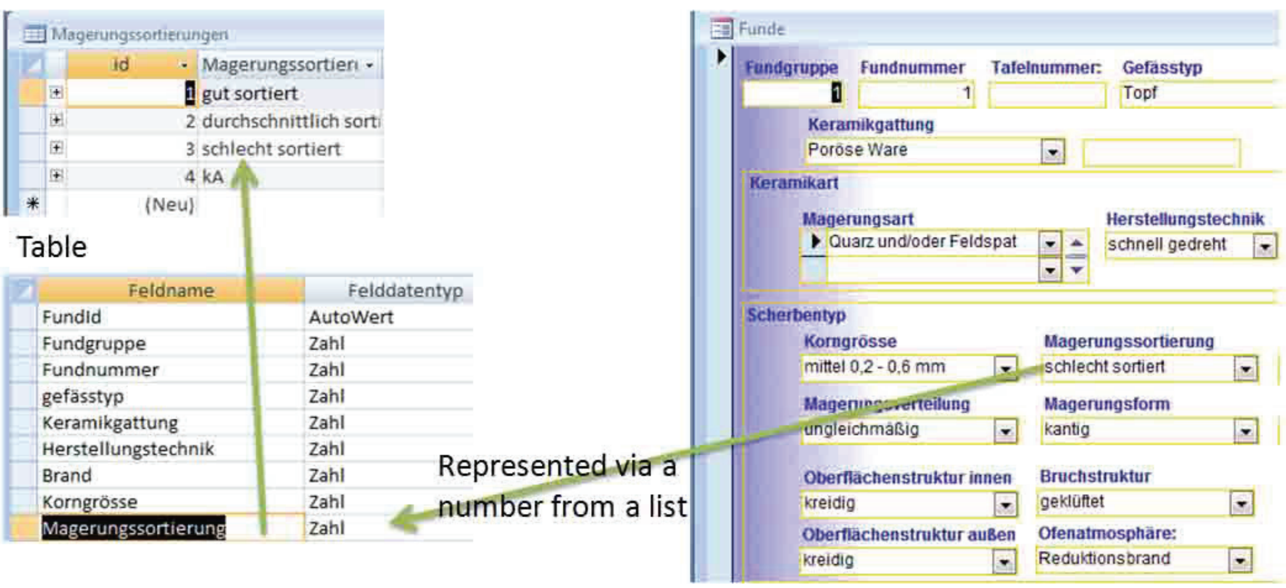

Fig. 1. Attribute values represented via lists.

[1] Just to name some possibilities: onset of aerial photo, photo interpretation, and stereo-photography.

description of metal or wood pieces would not be so detailed.

3.2. Different description lists

As a consequence of different data structures, each project uses its own lists to describe the records. A list stores all the data which could be selected from – for example a dropdown field or selection boxes. Such lists are often used to describe attributes like record type, sherd type, material, hardness and others. *Fig. 1.* shows the lists system.

The design and content of these lists depends – analogous to the data structure – on the different record types. The lists assure that the data is structured within an application. Based on such lists it is easy to run queries, save time during data collection and to avoid literal errors. But this (desirable) structuring will be obsolete if the content is provided to other projects because each project uses its own set of lists. Consequently, such lists are not useable for cross project data exchange unless further transformations are applied.

3.3. Different language

The aspired cross linking between projects raises the issue of different languages. The majority of records are stored in the mother tongue of the project group. This impedes automatic queries and reporting.

3.4. Different accessibility

Depending on the technical solution and the system architecture, the accessibility of the systems differs. Some applications run on local computers and therefore cannot be accessed by others. If other persons need access, the records have to be copied. If one side makes some changes (e.g. a project member) they do not affect the source. Thus, collaboration is very difficult (F. Eckkrammer *et al.* 2007). A big threat to such single user solutions could be data loss, for example caused by a hardware defect, a virus or burglary.

Some advanced systems are already connected to the internet and are accessible no matter where the user is located. Such software designs allow collaborative work within the project team. However, few systems contain interfaces for data exchange to external applications or possibilities to abstract data or to tag them with security constraints.

The widely used system architecture of isolated applications builds the basic problem addressed by this paper.

3.5. Lack of knowledge about relevant DB's

Many groups of researchers are working in similar fields and deal with similar archaeological questions. Therefore they are in need for the same information (=data). Because of these isolated systems, archaeologists have to deal with the problem that only a small group can access or even knows about this particular information.

A discussion of several approaches to solution follows the listing of the core problems. As pointed out in this chapter, the problems intertwine. The approaches are listed so as to make increasing cross linking and interoperability possible.

4. Description of approaches

4.1. Different database systems and data-structure

To solve this issue two different methods were tested:

1) Standardisation
2) Data mapping

Ad 1)
The efforts to create a set of standard attributes cross-cutting periods and cultures result in a very large and inflexible set of attributes. An additional drawback was that this new format only addresses new projects and does not provide a solution for existing ones. Consequently the intention of standardization was subordinated to the second approach.

Ad 2)
During the mapping process databases are wrapped by an interface. The interface defines a mapping from the local attributes to the attributes of the interface. The interface uses a standard like the CIDOC CRM model (ISO 21127 standard). This process is called mapping. Several tools for mapping data to the CIDOC CRM model are already available. Additionally the interface has layers where security, data sharing and user settings can be taken into account.

Fig. 2. shows the concept of data sharing/exchange between two different databases via two locally installed interfaces. The data is submitted

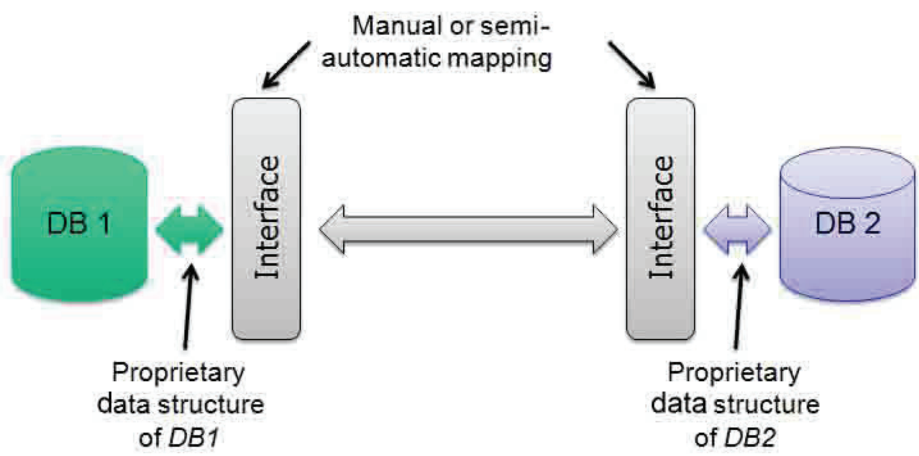

Fig. 2. Data exchange via the interface.

from one interface to the other in a standardized format. The receiving system (which has also mapped the attributes to the interface) receives the data and maps it back to the local/ own format. Also linking of data between databases as well as change notifications would be supported.

However, the mentioned approach did not address the discovery of relevant database systems.

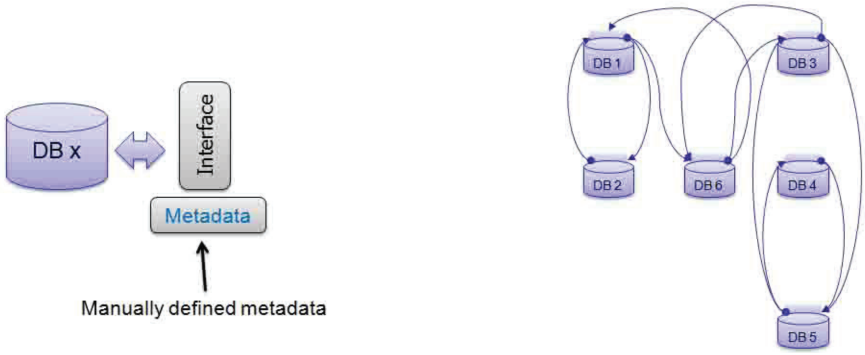

Fig. 3. Data exchange via a peer to peer concept and adopted by metadata.

Therefore the concept would be extended with a set of metadata. Those metadata describe the content of the database/system in a structured way so that it could be processed automatically.

The metadata contain, among other entries, data like:

– LanguageCode
– Connection strings
– Custodian
– Licence Types
– Sitenames
– Site Location
– Categories

To distribute the metadata, two methods are supported:

– Centralized metadata Server
– Distributed /peer-to-peer concept

For the first option each database submits the metadata to a centralized server. This server can handle search requests and provide the information to the clients.

The second approach works like a peer-to-peer system. Each database collects the metadata collection from the requesting database and stores it locally. After requested by another these two systems exchange all metadata as well. The concept is shown in the *Fig. 3*.

The next step should provide a solution for the usage of different lists to describe properties of a record as well as the usage of different languages. To solve this issue an approach from the eHealth sector was adapted for the archaeologists.

The lists will be identified via OID numbers which are worldwide unique identification numbers. Each institution or person can get such numbers. This OID and some other attributes could be used to describe the lists in a unique and assignable way. The identification string could be fragmented so that a wide range of sub-groups would be possible.

The following attributes are used to describe a list:

– codeSystem
– codeSystemName
– code
– displayName
– codeSystemVersion

This information would be added to the records during the exporting or sharing process. Other databases could easily identify the used description lists.

The example shown below describes a shape/the usage of a finding:

- codeSystem: 1.3.6.1.4.1.32.12176.1
- codeSystemName: pottery description List of Celtic Museum Hallein (Dürrnberg)
- codeSystemVersion: V1.0.10
- code: 207
- displayName bowl

Thereby the **codeSystem** describes the list which is used to describe the finding by the OID number. The **codeSystemName** describes the list in a human readable manner (the name of the list). The usage of a **codeSystemVersion** allows updating the list. The **code** attribute describes the selected entry (row) of the list by a unique number (unique number inside the list). And the **displayName** describes the selected list entry in a human readable manner.

Due to the usage of a centralized registry the lists could be translated (semi-)automatically by the server. Hence the multi-language issue is also addressed. Of course not all description attributes are entered via lists, but the translation of the lists is at least a beginning. So cross language data exchange would be possible. The metadata entries could also be translated automatically if centrally stored.

5. Conclusion

In this paper we have described approaches to some core problems which have to be solved in order to facilitate the cross linking of currently isolated databases which would make project crossing queries and automated searching possible. In a nutshell the objective is to create a holistic way for capturing, managing and releasing data of excavations under consideration of the knowledge of other projects as well as the exchange between them. We have demonstrated the advantage of the approaches and of the cross linking by several examples. The benefit of providing consistently structured data over different projects and the fact that this approach is meant to use available databases without editing existing databases itself highlights the genuine property of this approach.

6. Perspective

Recognizing the larger pattern in the development of human culture it appears more important to establish connections between single excavations, projects, or museums. That is the way to evince distribution of goods[2] and not just pointing to certain contacts due to imports. To name an example for clarification, questions for a tomb could be:

- Where is it located?
- How is it arranged?
- Which period can be observed?

Another topic could be comparisons of intangible culture as language and signs[3] leading to a better understanding of archetypes or coincidental parallels between objects.

Different viewing angles are possible: On the one hand it is essential to couple all questions as chronological or regional parallels concerning objects or structures. On the other hand – meant in literal sense – an observation of an excavation site can be made either from above (beginning with the position of the site down until the object) or from the side, allowing the interpretation of stratigraphy, offering a multi-layered and complex sequence of dynamics due to innumerable events.[4]

Thanks to technical development, a lot is possible nowadays. But a difficult step is convincing members to exchange the data with others. The more members (of more different disciplines) join, the better the global research.

Descriptions of finds are still far away from being homogenous. Nevertheless all of them could be embedded or integrated for a more elaborated and detailed typology. There are still some uncertainties. For example the problem, how finds from different regions and epochs can be described in a standardized way. Simplifying the handling of a large amount of finds during the creation of typologies/definitions and/or catalogues is important, also as *criterion*

[2] That knowledge also allows interpretations regarding transfers of intangible assets as knowledge, ideas and religion.

[3] There are already first steps working on sign formation and lexical ontologies in a cross-cultural perspective.

[4] Each point of view – be it objects or structures – consists of data, photographs, drawings with measurement, or integrated geo-data and can be examined separately. Through cross-references it is possible to associate a find to a region, a culture, an excavation, or whatever is necessary for the research. Interactive maps (as satellite photos, morphological plans, etc.) with either an entire overview or within a chronological frame, can be linked to each other.

for scientists whose pursuit is generally outside of certain considerations. This is important to bring more transparency and to make scientists able to communicate and discuss with others beyond their specialization.

The opportunities of being interlinked throughout the world are not yet used in some sciences on a scale as they should be.

References

Droysen, Johann Gustav (1993[9]). *Historik, Vor lesungen über Enzyclopädie und Methodologie der Geschichte*. München.

Eckkrammer, Florian, Tobias Eckkrammer and Rainer Feldbacher (2008). Possibilities to exchange, map, and administer finds from different archaeological projects within databases. *Kulturelles Erbe und neue Technologien. Workshop 12 – Archäologie und Computer*. Wien.

Leclant, Jean (ed.) (1996). *Archeologia. Percorsi Virtuali nelle Civilta´ Scomparse*. Milano: Mondadori.