Concept for an Ontology Based Web GIS Information System for HiMAT

Gerald Hiebel – Klaus Hanke

University of Innsbruck
Surveying and Geoinformation Unit
{gerald.hiebel; klaus.hanke}@uibk.ac.at

Abstract

The multidisciplinary research program HiMAT investigates the History of Mining Activities in the Tyrol and adjacent areas with the goal to answer archaeological, cultural, environmental and social questions related to historic and prehistoric mining activities.

A web based spatial information system (web GIS) should enable integration, communication, retrieval and presentation of data within the project. The main challenge is the diversity of intellectual concepts, research methods and therefore heterogeneous data coming from 10 participating disciplines. Selected parts of the formal ontology of the CIDOC CRM were chosen for knowledge representation.

Within the project a document management system is used to store all kinds of relevant digital data. A geodatabase is used to store spatial information and metadata while a web GIS provides functionalities for visualization, spatial selection, the relation of spatial information to other information and access to documents stored in the document management system. A simplified system architecture shows implemented components and open issues.

Keywords

web GIS, database, ontology, cultural heritage, CIDOC CRM

1. Introduction

In March 2007 the multidisciplinary research program HiMAT started at the University of Innsbruck to investigate the History of Mining Activities in the Tyrol and adjacent areas. One of the tasks of the Surveying and Geoinformation Unit is to build a web based spatial information system (web GIS) to enable integration, communication, retrieval and presentation of data within the project HiMAT and to the public. In order to do this we encountered the need for a concept of knowledge representation. The information available in HiMAT should be represented using metadata and parts of the formal ontology of the CIDOC CRM serve as data model. A document management system exists within HiMAT and information stored there should made accessible through a web GIS if they relate to a location. Moreover data of the document management system should be integrated in the concept for knowledge representation. Spatial information and metadata will be stored in a geodatabase. Web GIS functionalities for visualization and data management access this geodatabase. With the help of the parts of a simplified system architecture we want to show existing components and address open issues.

2. HiMAT – History of Mining Activities in the Tyrol and adjacent areas

The multidisciplinary research program HiMAT started in 2007 and will last for 10 years. It is situated at the University of Innsbruck with partners in Germany and Switzerland. The period of investigation ranges from prehistory to modern times.

10 disciplines (Archaeology, History, Linguistics, Surveying and Geoinformation, European Ethnology, Botany, Archeozoology, Dendrochronology, Petrology, Archeometallurgy) organised in different project parts with a major focus on archaeology are involved in the research on archaeological, cultural, environmental and social questions related to historic and prehistoric mining activities (Hanke 2007). In the first four years there are four key areas of research located in the western part of Austria as shown in *Fig. 1*. Schwaz in Tyrol was one of the centers of silver and copper mining in medieval times and the Mitterberg must have ranged within the most important copper producers during the Bronze Age. This multidisciplinary project is sponsored by the Austrian Science Fund (FWF).
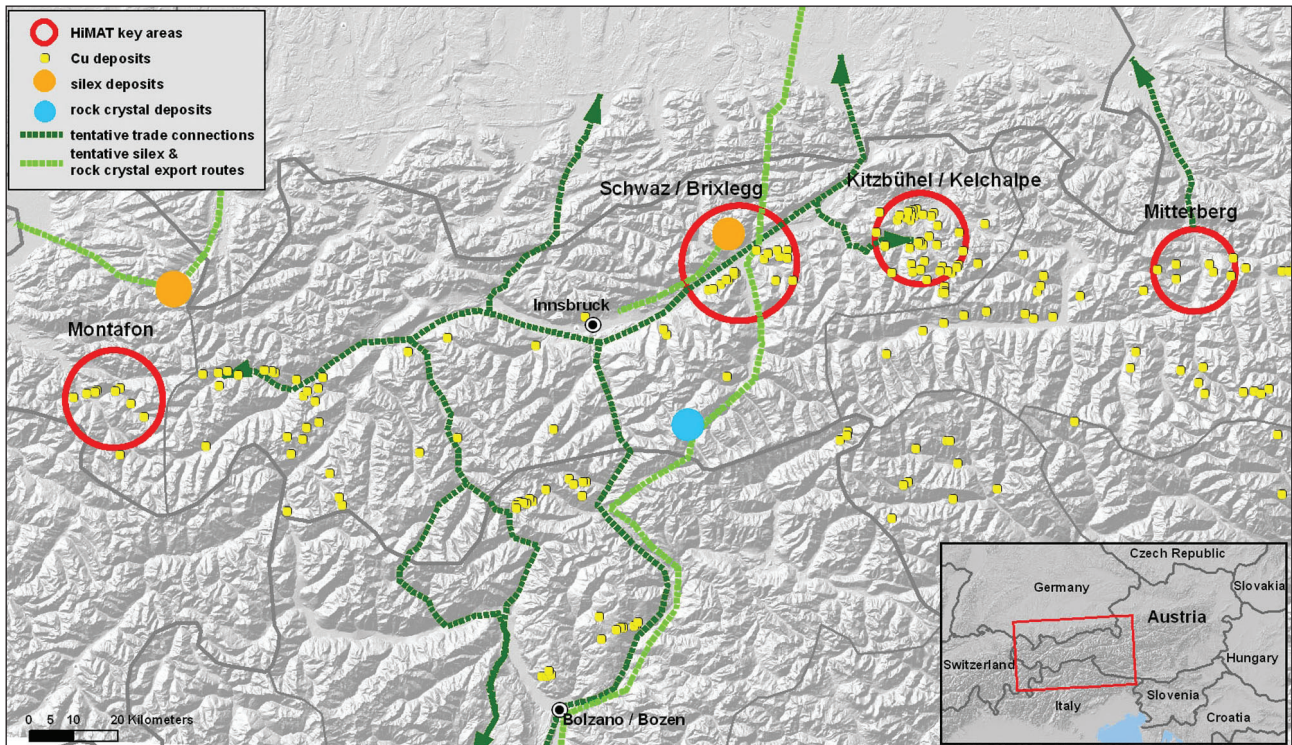
Fig. 1. Key areas of research.

3. Knowledge representation

With all this disciplines participating in the project we face a diversity of intellectual concepts, research methods and therefore heterogeneous data. After a process of investigation on the digital and analog resources of the different project parts it became obvious that it is not possible to organize and handle all data and information of the project parts in detail. Moreover there had to be choices made which data is relevant for the whole project. For this reason we decided to build up a system of metadata that gives an overview of the available information and links to data and providers. The first step is to decide what information should be exchanged between the project parts. Workshops with the participating disciplines lead to a list of fields that should contain the relevant information.

A common concept of knowledge representation within the whole project is essential in order to have a shared understanding of the respective metadata field. One important part in such a concept is to assure the integrity of information meaning that different entities should be clearly separated and only stored once. A second part is to define relations of the fields to each other. For these two

requirements a data model is necessary. Given the complexity of the subject an object oriented semantic model is a state of the art choice. The formal ontology of the CIDOC CRM (Crofts *et al.* 2007) provides the necessary semantics and is an ISO standard that starts to be frequently used in archaeology.

So we chose this ontology as a base concept and the next step was to define which classes of CIDOC CRM are necessary to represent the metadata information we agreed upon. *Fig. 2* shows the main CIDOC CRM classes used in HiMAT (with their notion of Exx) in a diagram. *Table 1* describes their
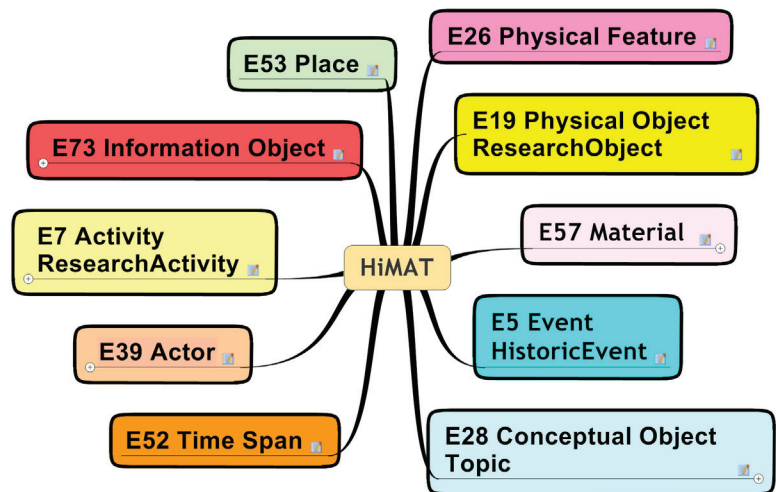

Fig. 2. CIDOC CRM classes used in HiMAT.

CIDOC CRM class name Subclasses or related classes	HiMAT Scope	Examples

E26 Physical Feature
   E25 Man-Made Feature
E19 Physical Object
   E22 Man-Made Object
   E20 Biological Object
	research object
	   artifacts
		topography (hill), geology(copper deposit)
		   mine, settlement, copper deposit
		research Sample
		   tools, ceramics
		   trees

E 57 Material		copper, wood, bronze

E73 Information Object
   E38 Image
   E33 Linguistic Object
		PDF document, digital terrain model
		   (aerial) photo
		   interview, legend

E7 Activity
   E16 Measurement
	research activity	excavation
		   survey

E5 Event
   E7 Activity
	historic event
	   historic activity
		climate change, avalanche
		   mining, agriculture,

E52 Time-Span E49 Time Appellation		Bronze Age

E39 Actor
   E21 Person
   E 74 Group
		member of a project part
		project part, miner, peasant

E53 Place
   E44 Place Appellation
      E48 Place Name
      E42 Object Identifier
      E47 Spatial Coordinates
		Tyrol, Schwaz
		
		13,2323 15,54534

E28 Conceptual Object	topic	research discipline, organization, religion

Table 1. Details of CIDOC CRM classes used in HiMAT with examples.

scope within the project and lists subclasses or related classes that are necessary or useful. Examples from the project parts are given. The distinction of a historic event and a contemporary research activity was made as proposed by Cripps et al (Cripps *et al.* 2004).

After selecting the relevant classes the following task is finding out the relations (or properties with their notion of Pxx in CIDOC CRM) between them shown in *Fig. 3*. Decisions have to be made regarding the essential relations.
Examples:

A mining site like a tunnel or a gallery (E25 Man-Made Feature) was used for (P19) mining (E7 Historic Activity).

An excavation (E7 Research Activity) took place (P7) at Kiechlberg (E53 Place).

A hammer (E22 Man-Made Object) was measured (P39) with metallography (E16 measurement).

A fundamental process in order to enable integrated knowledge representation is building up a thesaurus of terms or keywords to be used within the project. These terms can be represented in CIDOC CRM as E55 Type. The HiMAT classes of the CIDOC

CRM were the starting point for a hierarchical structure. Within this effort the members of the project parts got to know the concepts of the CIDOC CRM. These keywords serve also as examples for the classes used in HiMAT and therefore provide a better understanding of their scope, what is essential when it comes to data entry. Up till now over 600 terms have been collected and agreed upon. Our aim was to organize them in not more than 3–4 hierarchical levels below the main classes, to facilitate data entry. *Fig. 4* shows the terms collected for "E7 Activity – Research Activity" which is one of the smaller parts of the Thesaurus. Existing CIDOC subclasses were used as terms (e.g.:E16 Measurement).

4. Document management system

Within the project HiMAT a large number of digital resources are gathered or produced by the project parts. A document management system (MS Sharepoint) provides the infrastructure to handle these data like PDF or MS Word documents, photos, maps, interviews, excel sheets or even databases.
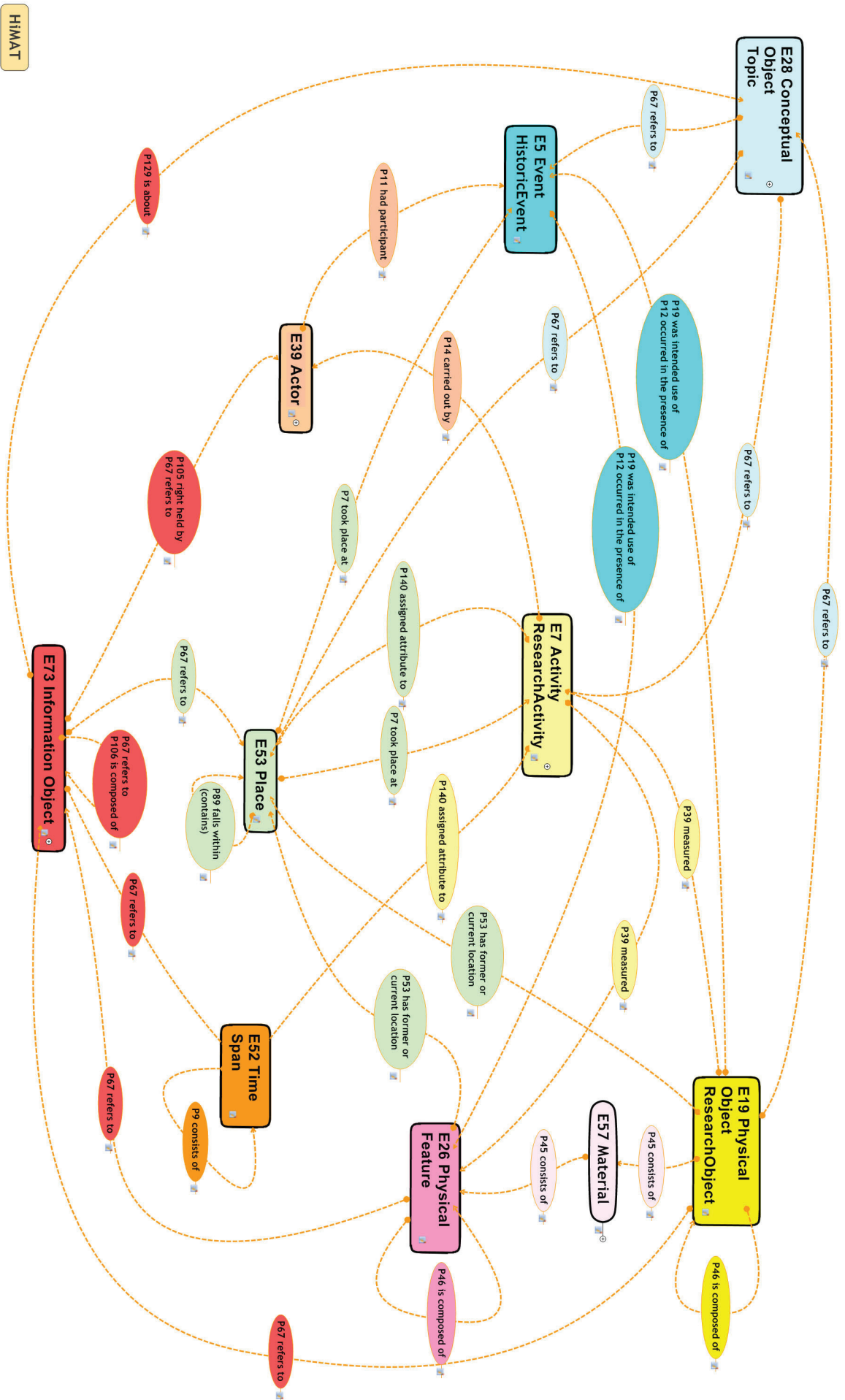
Fig. 3. CIDOC CRM classes (Exx) and relations or properties (Pxx) relevant for HiMAT.
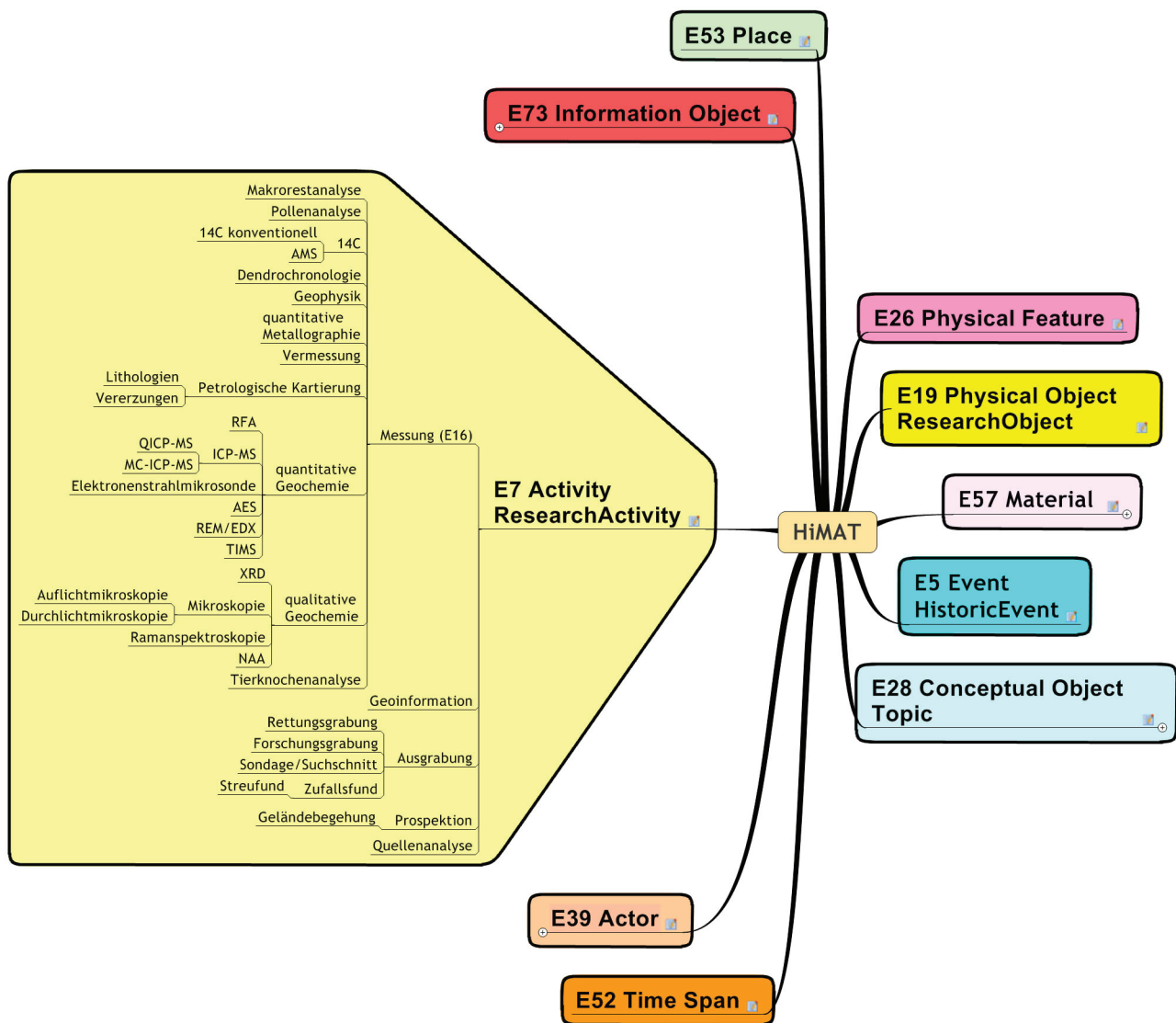
Fig. 4. Terms (in German) collected for 'E7 Activity – Research Activity'.

This type of digital resources corresponds to the class E73 Information Object in CIDOC CRM. A different kind of information stored within the document management system is handled by calendar functionalities. Project parts enter their research activities (E7 Activity) in MS Sharepoint calendars storing information of participating persons or groups (E39 Actor).

Our goal is to link digital resources and calendar entries for research activities with topographic names (E48 Place Name). That should be supported by spatial selection functionalities of a web GIS. Other metadata based on HiMAT CIDOC CRM classes ought to be related as well.

5. Spatial information – geodatabase

Part of the metadata dealing with the CIDOC CRM class E53 Place contains spatial coordinates (E47).

As we want to use web GIS to access the metadata a particular focus has to be put on the storage and handling of spatial objects. The definition of the E53 Place object is leading the way.

E53 PLACE: This class comprises extents in space, in particular on the surface of the earth, in the pure sense of physics: independent from temporal phenomena and matter. (Crofts *et al.* 2007)

This means that an instance of E53 Place is neither identical with a given name for that place nor with a settlement, a mine, a copper deposit, an object found there. Names of places can change over time, settlements can vanish and excavations have limited duration. Therefore it is important to store the instance of a place only once and all other related things such as names, settlements, mines, copper deposits, objects or excavations associated with it.

To enable the integration of coordinate information necessary for a web GIS with other metadata they should be stored in the same database, a geodatabase that has to be able to store spatial objects and should be available to Desktop GIS users to do advanced geoprocessing and analysis. A geodatabase that can be accessed by various applications is needed as described by Brinkhoff (Brinkhoff 2005).

Not only is the structure of a database essential for its usefulness but moreover all its contents. If data entry is not standardized it will not be possible to identify features of the same kind. The classes of the semantic model already give a rough distinction but the detail information is provided by the terms of the thesaurus.

Another kind of spatial information used in GIS are different types of raster data (maps, ortho images, digital terrain models …). In our model they would correspond to E73 Information Objects and should be stored in the document management system. Due to the size of these data and their accessibility for the web GIS it is necessary to store them in a file system or a geodatabase, depending on the raster functionalities of the product. As the manner and location of storage of raster data is a critical factor for the performance of a web GIS this question has to be evaluated thoroughly.

6. Web GIS functionalities

A web GIS can have different functionalities but its main purpose is visualization. Information with a spatial relation should be displayed. The idea is to use CIDOC CRM classes related to place as folders for layers representing different types of these classes as demonstrated in *Fig. 5*.

Examples:
CIDOC CRM class as folder: E26 Physical Feature
Type as layer: copper deposits, mines, settlements
CIDOC CRM class: E7 Activity – research activity
Type as layer: excavations, surveys
CIDOC CRM class: E73 Information Object
Type as layer: maps, digital terrain models, PDF documents, photos

It should be possible to apply queries and filters to the layers to retrieve only elements for example concerning the Bronze Age.

To enter information with a spatial relation a web GIS needs certain functionalities which are not commonly found in these applications because they go beyond visualization. If a location (E53 Place object) that should be related to information already exists within the system there should be the possibility to graphically select that location as shown in *Fig. 6*.

If the location does not exist it must be possible to enter a new location either by input of coordinates or by graphically entering a new point with the necessary attributes for an E53 Place object like E48 Place Appellation.

For the concept of place we use it is very important to check if an E53 Place object already exists at the
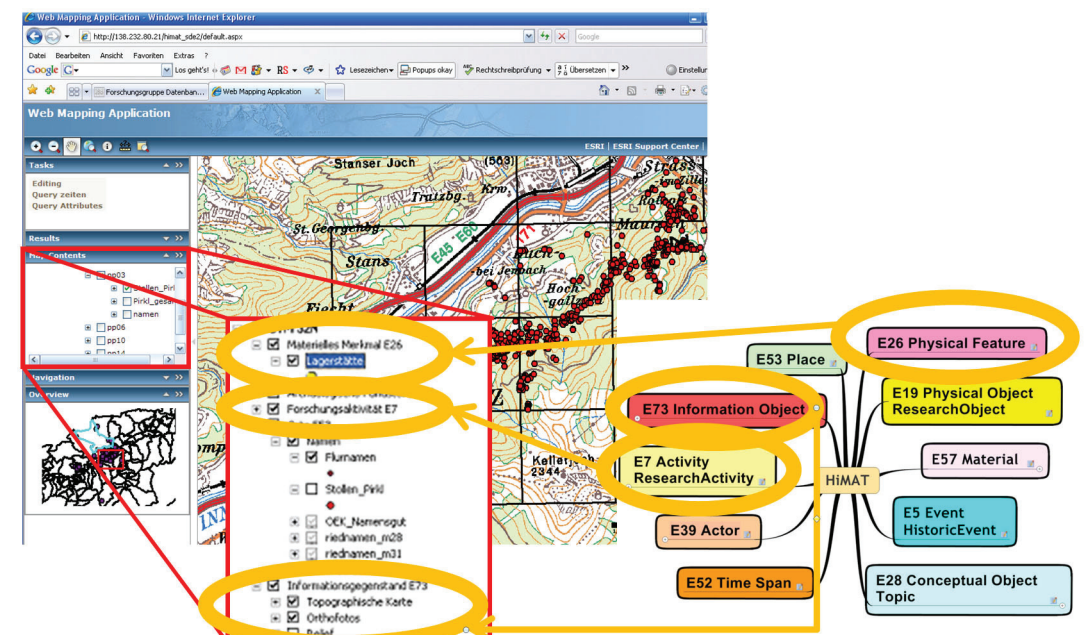

Fig. 5. Visualization of information with a spatial relation.

location you want to attach information to, in order not to enter a location twice.

The next step is to relate the selected or entered location (E53 Place) to other information like E26 Physical Feature (settlement), E7 Research Activity

Fig. 6. Graphical selection of an existing location.

(excavation) or E73 Information Object (PDF document) as illustrated in *Fig. 7*. Information on these objects has to be entered first.

Another functionality of the web GIS is the access to documents stored in the document management system (DMS). Places that have been related to documents with the process described before have an URL (e.g.:https://sp01.uibk.ac.at:10122/himat/pp14/mauken.pdf) provided by the DMS (*Fig. 8*). Security restrictions implemented in the DMS regulate the access to the files.

7. System architecture – open issues

With the help of the parts of a simplified system architecture we want to show existing components and address open issues.
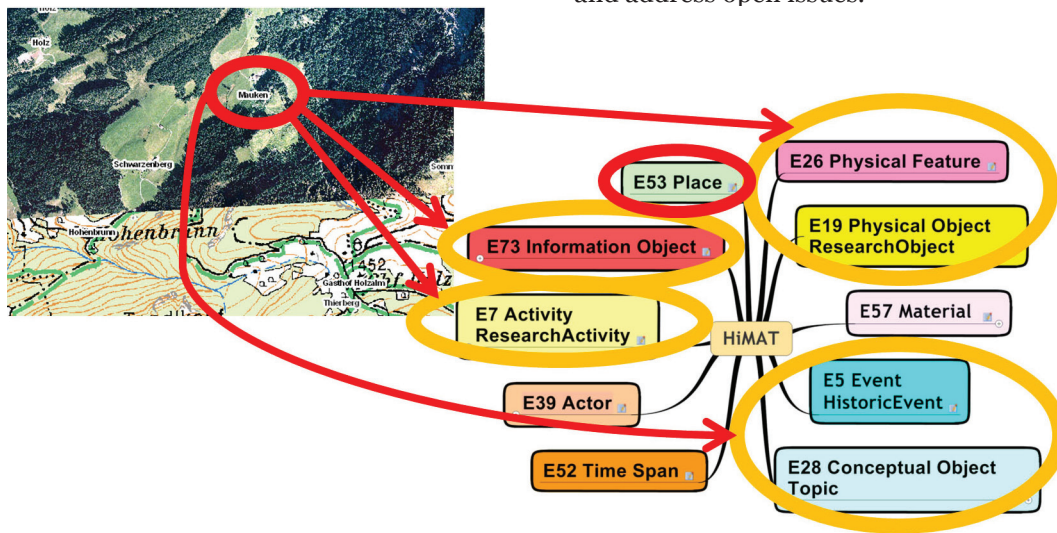

Fig. 7. Relate a selected location (E53 Place) to other classes.
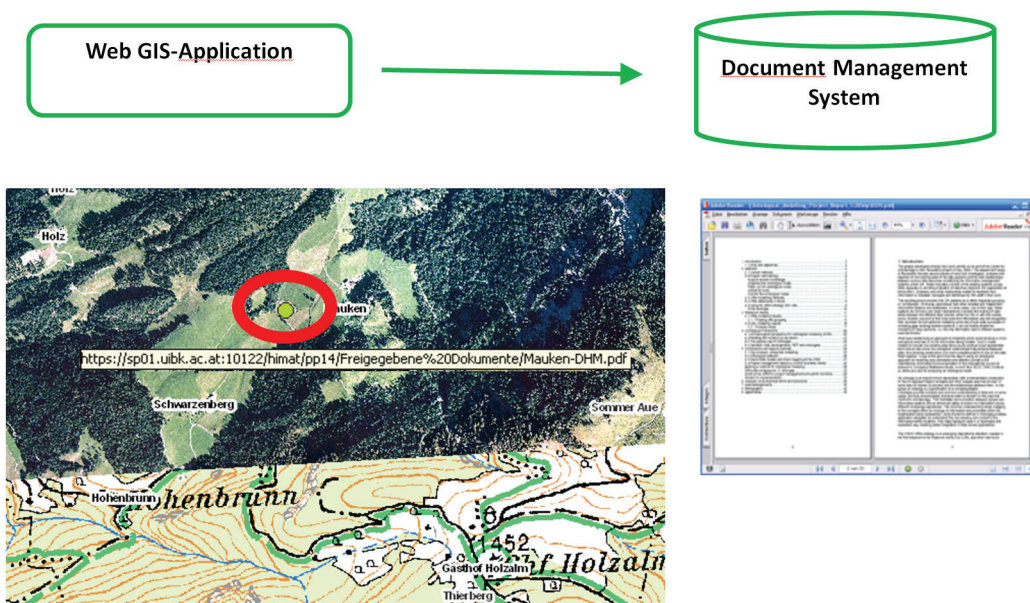

Fig. 8. Web GIS access to documents stored in the document management system.

In our view the system should be built upon a geodatabase holding spatial objects, metadata and relations between them. The database should be open to be used with different software products. This is the central unit of the system and the other parts of the system access the information stored in the geodatabase or have an interface to exchange information. The web GIS retrieves information for visualization purposes and offers functionalities to enter spatial and non-spatial information including their relations.

Project parts databases and the document management system are external databases that have to be linked with interfaces to the central database.

Fig. 9 shows all components of the proposed system. Right now the existing components are a web GIS prototype for visualization, a document management system and the respective databases of the project parts. Open issues are the geodatabase holding spatial objects, metadata and their relations. Our object is the implementation of the data model in a relational database. Interfaces to the project parts databases and to the document management system are still challenging tasks to be accomplished. Software solutions to enable web GIS functionalities of location selection or input and the relation of locations to other information are currently worked on and tested.
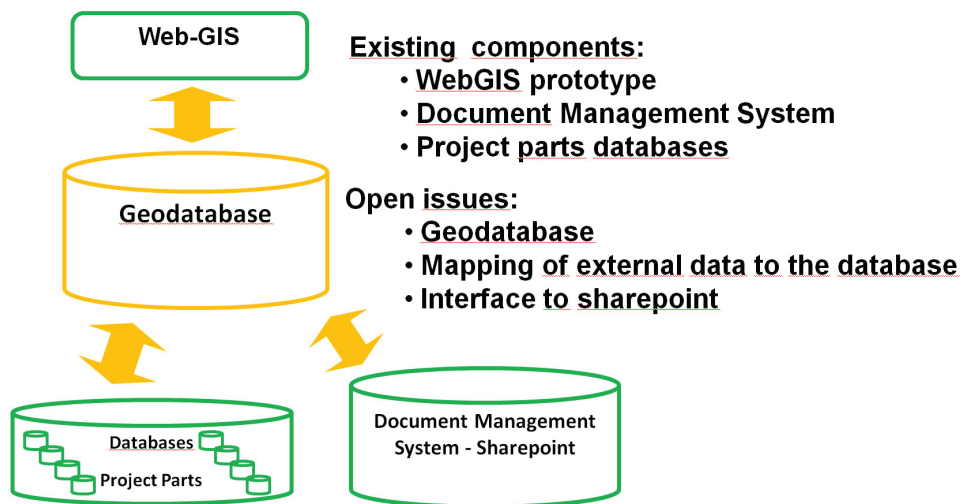
References

Brinkhoff, Thomas (2005). *Geodatenbanksysteme in Theorie und Praxis*. Heidelberg: Wichmann.

Cripps, Paul, Anne Greenhalgh, Dave Fellows, Keith May and David Robinson (2004). Ontological Modelling of the work of the Centre for Archaeology *CIDOC CRM Mappings, Specializations and Data Examples* (http://cidoc.ics.forth.gr/crm_mappings.html)

Crofts,Nick, Martin Doerr, Tony Gill, Stephen Stead and Matthew Stiff (eds.) (2007). Definition of the CIDOC Conceptual Reference Model Version 4.2 *Official Release of the CIDOC CRM* (http://cidoc.ics.forth.gr/official_release_cidoc.html)

Hanke, Klaus (2007) Contribution of Laser Scanning, Photogrammetry and GIS to an interdisciplinary Special Research Program on the History of Mining Activities (SFB HiMAT). *The ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences Vol. XXXVI – 5/C53* (ISSN 1682 - 1750) and *The CIPA International Archives for Documentation of Cultural Heritage Vol. XXI – 2007* (ISSN 0256 - 1840).

Fig. 9. Simplified system architecture.