OpenATLAS

An Open Source Database Application for Archaeological, Historical, and Spatial Data

Stefan EICHERT

University of Vienna, Department of Prehistoric and Historical Archaeology

Abstract: OpenATLAS is an open source database application for the work with archaeological, historical, and spatial data. As a backend it uses PostgreSQL with PostGIS in a server based version. Alternatively it can use a file based, locally installed backend with Sqlite and Spatialite.

The backend fulfills state of the art technical standards and is compatible to nearly all GIS-applications like for example Quantum GIS.

Regarding the model of the data it uses classes and properties from the CIDOC-CRM, which provides a high compatibility and sustainability of the recorded information.

The graphical frontend aims at being as user friendly as possible and reducing the complexity to an extend necessary. It offers a standardized workflow for data acquisition and tools for managing, filtering, searching and exporting the data.

OpenATLAS' open structure and the object-oriented model furthermore provide the possibility to adapt it to certain individual requirements.

Keywords: Database application, Open Source, CIDOC-CRM, GIS, Archaeological Data

Fig. 1 – The OpenATLAS Logo (Design: Gregor Brundke/Stefan Eichert)

Introduction

This paper presents the OpenATLAS project (www.openatlas.eu) and discusses its technical and conceptual background. OpenATLAS aims at providing a user-friendly database application for the work with archeological, historical, and spatial data. It bridges a gap between the needs of scholars from the field of humanities for a tool to handle their data and the requirements of superior institutions and the overall scientific community for certain data standards and sustainability (e.g. THORNES & BOLD 1998). The development is carried out by a small team from the University of Vienna. It is funded in the framework of the

Project "Eastern Alps Revisited" (FWF – Austrian Science Fund P. No. P24045) of the Austrian Academy of Sciences – Institute for Medieval Research and the University of Vienna, Department of Prehistoric and Historical Archaeology. It is furthermore supported by the University's Faculty of Historical and Cultural Studies. Further cooperation is held with the Research Group "Digital Humanities and Digital Cultural heritage" of the "Cluster of Excellence – Asia and Europe in a Global Context" of the University Heidelberg and the "Römisch Germanisches Zentralmuseum" in Mainz/Germany.

Features

Data acquisition

OpenATLAS offers a workflow for the standardized acquisition of archaeological and historical data with all relevant information. One can digitize information from archaeological publications, excavations' documentation, historical sources like medieval charters, bibliographic details and many more. The graphical user interface (fig. 2) automatically links all the necessary relations between the various entities and their sources. A digital import of data is possible as well.

Data Management

OpenATLAS can be used to manage large databases. Users can browse data, display relations between various entities, query data, edit entries and much more. We designed it to handle a large amount of archeological sites, finds and their connections.

Analyses, Reports and Visualization

Once the data is collected in OpenATLAS, one can easily use it for further analysis. With PostgreSQL and PostGIS already a broad variety of statistical and spatial analyses is possible. All sorts of queries or reports can be generated and their results can be used in other applications like R, SPSS, Calc, Excel, in text editors like Word or Open Office Writer and GIS programs like ArcGIS, Qgis and many more for analyzes and visualizations.

Data Model

Although there have been a lot of attempts in the recent past to establish data standards to handle digital archeological information we are far away from a consistent solution that fulfills everyone's requirements. A general standard could in fact homogenize the data and allow an overall and common handling of various information from different sources. However, this would on the other hand foreclose individual approaches. Such general models often do not offer the possibility to cope with special requirements each project usually exhibits. If the models do, they are way too complex to be accepted and used by non-database-experts. On the one hand, this leads to unsatisfying solutions, where certain individual, but important information, are left aside to fulfill a general standard. On the other hand, we experience detailed and specialized, but isolated solutions that are not compatible to other projects and have a questionable sustainability (On the discussion of archaeological databases see for instance. LABRADOR 2012).
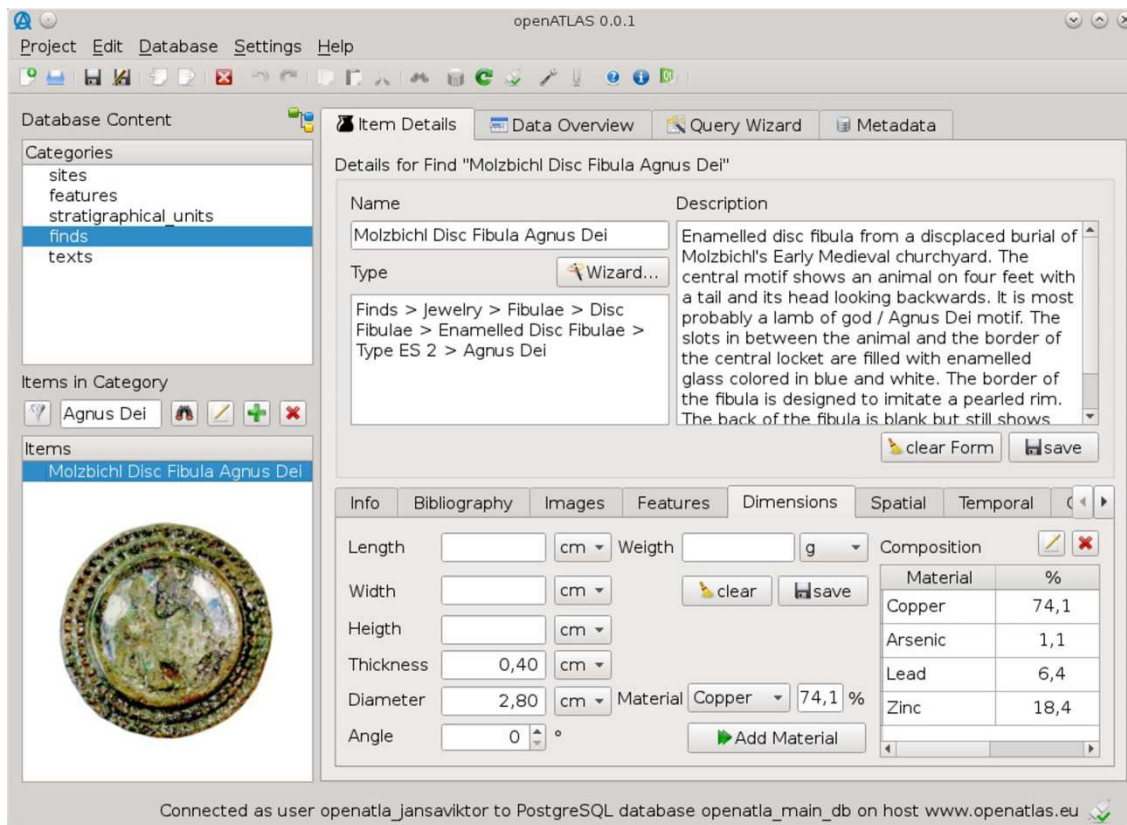
Fig. 2 – Screenshot of the user interface (Opensource Version, Design: Viktor Jansa)

One approach to offer a common system to deal with different types of data, especially from the field of cultural heritage, and at the same time allow individual modeling and mapping of information is the CIDOC-CRM (LE BOEUF et al. 2013). It seems to become more and more accepted as a common standard in digital humanities. One problem though is certainly the gap between potential users like archeologists or historians and the – on a first view – very complicated model.

Here comes OpenATLAS into play. It aims at providing an interface between these two parts. It uses classes and properties of the CIDOC-CRM to map its data and offers a graphical user-interface that reduces the complexity by automatically creating the necessary links between the entities that are investigated. At the current stage of development OpenATLAS can deal with 4 overall types of information (fig. 3):

1. Physical entities like archaeological sites, finds, stratigraphical units

2. Human entities like persons, groups of persons and institutions/organizations

3. Temporal entities like activities, events, phases, actions

4. Sources like articles, images and documents that refer to other entities

Each of the entities mentioned above may be related to other ones of the same or different class and also be subdivided into one or more subunits.
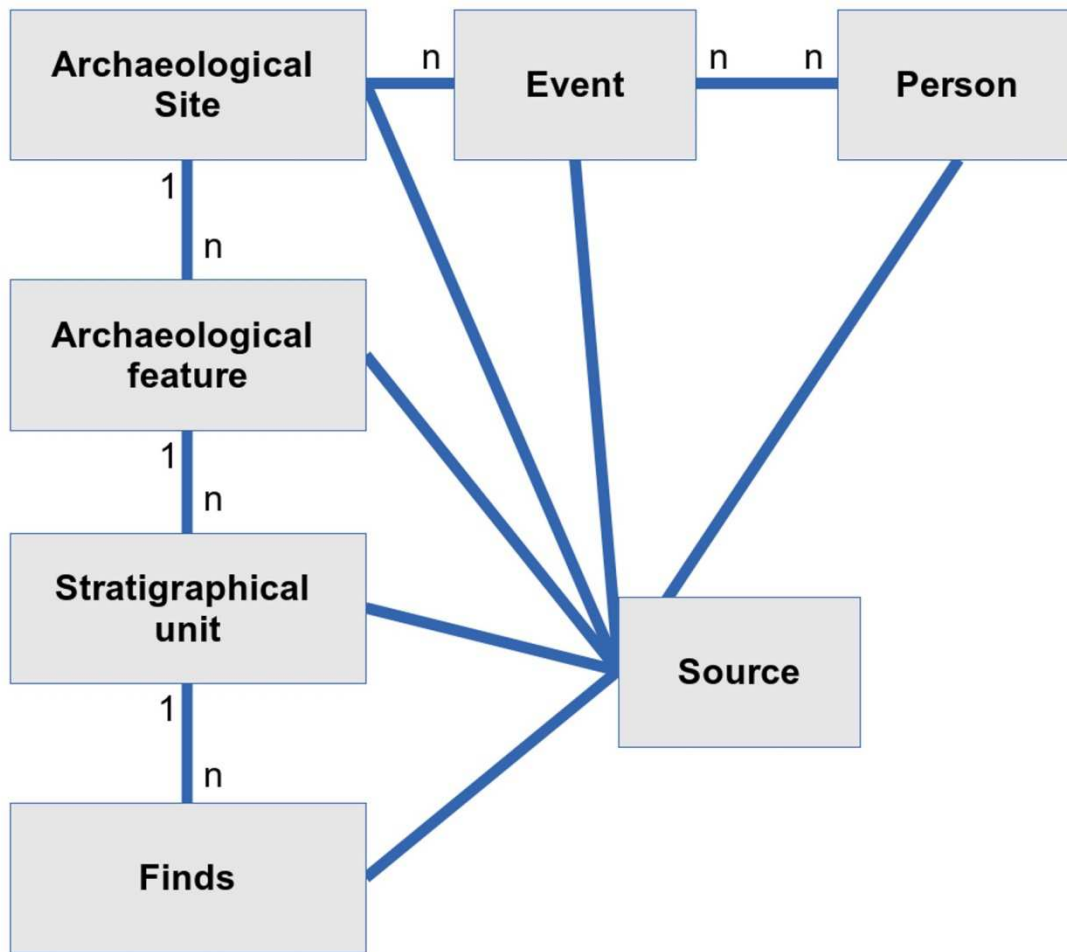
Fig. 3 – The general data model of OpenATLAS (Design: Stefan Eichert)

The focus of the current development lies on archaeological objects and thus OpenATLAS uses a 4-level-hierarchy to deal with archeological information and their relationships. On the top level there is the archeological site itself. It is defined as the whole physical object that contains all the other subunits that form the site. This site can e.g. be a settlement, a building, a cemetery, a cave, a road or simply the place where finds haven been found. An archaeological site can be composed of one or more archeological features, like graves (in case of a cemetery), walls, deposits etc. (in case of a building) or for example a conceptual unit like the unknown context of a strayfind. These features are again composed of subunits, or more precisely of their stratigraphical units that were documented during excavation or logically reconstructed e.g. from older literature. For a grave this may be for example the backfilling, a coffin and primary as well as secondary burials/skeletons. These stratigraphical units may contain associated archeological finds like the finds from a certain stratum or the grave goods belonging to a skeleton, which form the fourth level of the hierarchy. This structure as described above is mapped using the CIDOC-CRM. An archeological site in OpenATLAS is for example represented by a combination of entities of certain classes from the CIDOC-CRM (fig. 4). The nucleus is an entity of class E18 (physical object) that is connected to various spatial, temporal and typological entities via certain properties. To define this entity of class E18 as an archaeological site a connection to an entity of class E55 (type) is used. This type has to be (archeological) site (not to be mistaken for CIDOC-CRM class E27 "Site") or one of its subtypes.
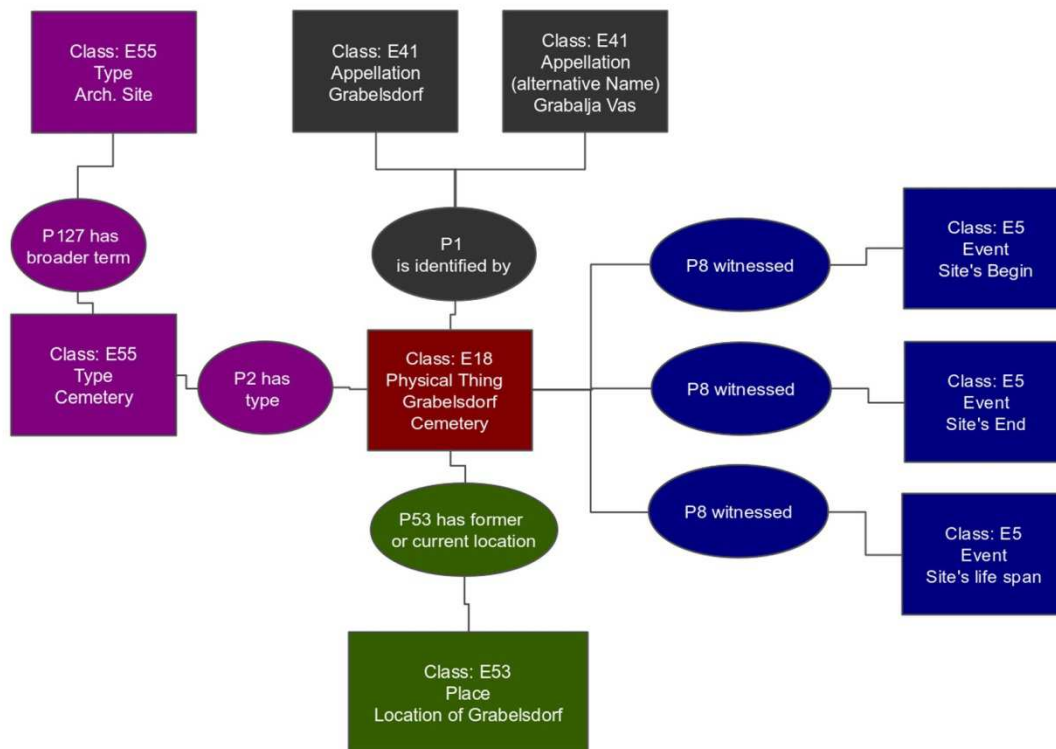
Fig. 4 – The core-mapping of an archaeological site in OpenATLAS (Design: Stefan Eichert)

The same holds for features, stratigraphical units and finds. They are mapped in the same way and their core is represented by an entity of class E18 (feature and stratigraphical unit) or E19 (finds). To define whether it is one of the mentioned archaeological categories a connection (P2 "has type") to an entity of class E55 (type) has to be given and this type has to be "feature", "stratigraphical unit" or "find" respectively be one of their subtypes. Subtypes of types are defined by Property P127 (has broader/narrower term). The hierarchy between different archeological units is recorded by the property P46 (is composed of/forms part of). A cemetery for example is composed of various graves that are again each composed of certain stratigraphical units that may contain certain finds.

This "lightweight" mapping of archeological information on the one hand aims to be comprehensible enough also for non-specialist users and on the other hand still offer the possibility to expand it with further classes and properties to model and map more complex situations and fulfill the requirements of experts.

Technical Background

OpenATLAS consists of a database backend and various graphical frontends.

Backend(s)

The backend is used to store the data and also takes care of basic routines and calculations, like for example the creation of certain IDs, keys, geometries and the conversion of coordinates etc.

Technically spoken there are two solutions provided: One is a file based Sqlite-database (www.sqlite.org) with Spatialite GIS (www.gaia-gis.it/gaia-sins) extension. This one is stored locally on the user's harddrive and can be accessed only from there. The other one is a PostgreSQL backend (www.postgresql.org) with PostGIS (http://postgis.net) extension. This server based solution offers a multi user environment and can be accessed remotely via the internet.

Both backends were developed by Stefan Eichert. They have the same functionality regarding the way OpenATLAS handles their data and use exactly the same data model and structure. Therefore the data can be transferred from one backend to the other without any loss of information. Slight differences in the Sql-syntax do not affect the user.

The GIS-extensions of both database systems are compatible with every common GIS-program and follow the specifications of the Open Geospatial Consortium – OGC (www.opengeospatial.org). They allow to record data of various geometries (points, linestrings and polygons) and connect them with the information on the archaeological entities. Both, PostgreSQL and Sqlite as well as their GIS-extensions, are free and open source software, widely used and steadily developed further.

Frontend(s)

The frontend is used to access, edit, manipulate, query and insert data into the database. Currently there is one stable version that has been developed by Stefan Eichert in Microsoft Access with VBA (Fig. 5), which connects to the backend via ODBC. Although this frontend itself is considered to be free software, it needs proprietary software to run, even if it can be used within a cost-free runtime environment. To solve this issue a complete free and open source frontend that works as a stand-alone program is currently developed by Viktor Jansa in C++ within the Qt framework for Win, Mac and Linux (fig. 2). It offers the same functionality as the other frontend and will be published under an open source license.

The graphical user interface of the frontend aims at reducing the complexity and making the work with sophisticated information structures easier for the user by providing a standardized and efficient workflow for the data acquisition. It offers various tools for querying, filtering and managing the information. Furthermore the data can be exported to other programs for further analyses or visualizations.

The interfaces described are mainly designed for the work regarding textual content. For dealing with spatial data, a GIS-frontend for Quantum GIS (www.qgis.org) has been developed by Stefan Eichert (fig. 6). Users can easily connect it to the backend and use the built in PostGIS and Spatialite functionality of this open source GIS program to handle the geographical and spatial information stored in OpenATLAS. Within this program one can for example create new records by simply drawing them on a map background or digitize certain features from an excavation's documentation directly into the database. Also manipulating, moving, querying or searching data by spatial criteria can be done this way.
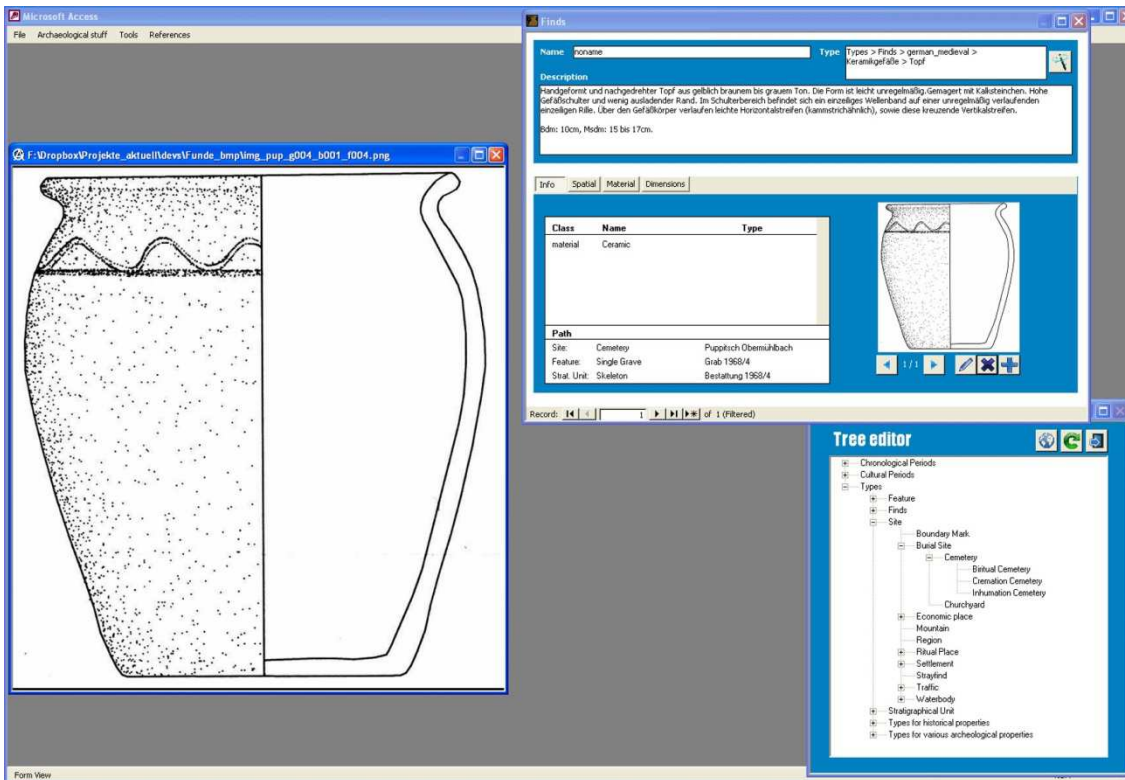
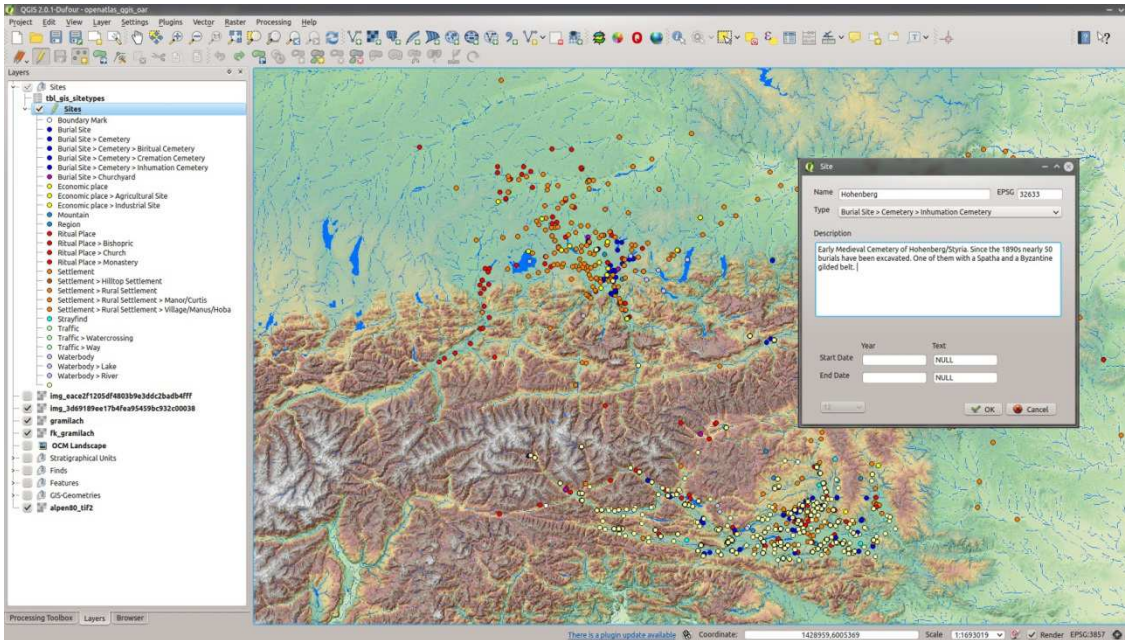Fig. 5 – Screenshot of the MS-Access frontend of OpenATLAS (Design: Stefan Eichert)

Fig. 6 – Screenshot of the Qgis frontend of OpenATLAS (Design: Stefan Eichert)

Synthesis and Perspectives

For the OpenATLAS Project it was very important to reach scholars, students and researchers that are not especially trained in IT or experts in databases as well as to also satisfy high-end users. One way to achieve this goal was the creation of a user-friendly interface. It guides the user through his work and reduces

complexity to a necessary extend. At the same time OpenATLAS' open structure and the design as an open source software allows further development and individual adaption. The use of the CIDOC-CRM offers the possibility to connect it to other projects that use the same reference model and provides a high compatibility and sustainability of the recorded data. The two backend versions (file- or server-based) and the possibility to seamlessly switch between them is another important aspect in this context. Many researchers prefer a local file-based install over a server-based one. They would therefore not accept such an application if it was too complicated and if the data were not physically on their computer. On the other hand a lot of projects with more than one co-worker need a multiuser environment and for them an exclusively file based version would not work.

Currently (2014) we are preparing the publishing of the first stable version as a stand-alone open-source program. It will be available on the project's website (www.openatlas.eu) as pre-configured install-packages for all common operating systems.

The source-code is hosted on GitHub (https://github.com/stefaneichert/openAtlas).

Currently there is also a "historical plugin" under development that will be able to deal with information from historical sources like charters or historiographies. From these sources further "entities" like persons, events or places will be "extracted" and their relationships to each other can be recorded in the database and mapped or visualized in various ways.

References

LABRADOR A. M. (2012) "Ontologies of the Future and Interfaces for All: Archaeological Databases for the Twenty-First Century" *Archaeologies* **8/3**: 236-249.

LE BOEUF P., DOERR M., ORE C. E., STEAD S. eds. (2013) *Definition of the CIDOC Conceptual Reference Model*. ICOM/CIDOC CRM Special Interest Group. http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.pdf

THORNES R. and BOLD J. eds. (1998) *Documenting the Cultural Heritage*. Getty Information Institute. http://archives.icom.museum/objectid/heritage/index.html