TOWARDS AN ARCHAEOLOGICAL INFORMATION SYSTEM:
IMPROVING THE CORE DATA MODEL

van Ruymbeke M. ¹, Carré C. ², Delfosse V. ³, Hallot P. ², Pfeiffer M. ⁴, Billen R. ³

¹ ULg Library, University of Liege, Belgium
² Geomatics Unit, Department of Geography, University of Liege, Belgium
³ Laboratory for User Cognition & Innovative Design, University of Liège, Belgium
⁴ Cluster of Excellence Asia and Europe in a Global Context, University of Heidelberg, Germany
mvanruymbeke@ulg.ac.be, Cyril.Carre@ulg.ac.be, Vincent.Delfosse@ulg.ac.be, p.hallot@ulg.ac.be,
michelle.pfeiffer@asia-europe.uni-heidelberg.de, rbillen@ulg.ac.be

Abstract
Geographical Information Systems demonstrated their utility in numerous archeological studies. However, such systems do not encompass all the characteristics of the archaeological data. Following an interdisciplinary approach, the Geomatics Unit (University of Liege) develops a robust, flexible and dynamic Archeological Information System. This research focusses on the general archeological data characteristics namely the spatio-temporal, heterogeneous, multimodal and imperfect components. In this paper, we present the latest outputs of this project and more specifically the integration of archeological data ambiguity. We depict the ambiguity management by introducing two new concepts: Lifemap and Interpretative sequence that make it possible to handle ambiguous hypothesis regarding historical objects within our system.

Keywords
Cultural Heritage, Archaeological Information System, Data model, Spatio-temporal database, Data Imperfection

1. Introduction
Since 2009, notably through the project Virtual Leodium, the Geomatics Unit of the University of Liège and their partners (Billen et al., 2009; Billen et al., 2012) work on the development of a robust, flexible and dynamic Archaeological Information System. Such system should provide solutions to the increasing needs of archaeologists and historians for a powerful set of tools for managing, archiving and exploring large archaeological data sets to encompass archeological problems (Burrough and McDonnell, 1998).

Due to the archaeological data complexity, namely its spatio-temporal, heterogeneous, multimodal and imperfect character, the development of an Archaeological Information System turns out to be a highly complex challenge that has been faced in the past resulting in numerous and versatile data models. Despite the high number of models, data imperfection is often ignored (Gonzalez-Perez, and Parcero-Oubina, 2011). Only few authors consider imperfection (namely for modelling purposes) (De Runz 2008, Pillot and Saligny 2011, Desachy 2012, Desjardin et al. 2012). These works distinguishes four different (but non-exhaustive) types of archaeological data imperfection: uncertainty, imprecision, ambiguity and incompleteness.

This paper presents the latest improvements of the Virtual Leodium core data model. In the previous archaeological information modeling works (Billen et al., 2012, Pfeiffer et al., 2013), we proposed solutions to handle geometrical ambiguity and incompleteness. Here, we present an improved model handling time and function’s imperfection. A new concept called Episode allows dealing with Events in a wider understanding than in the previous models. Moreover, new classes have been created. Interpretative Sequence organizes episodes into ordered path. Life Map amounts to Historical Object timeline, gathering all episodes even if they are contradictory. Finally, Agent and Figure classes complete the model.

The first part of the paper outlines the research context and our case study Virtual Leodium. We focus on the workflow, the archaeological information model and the current prototype of the archaeological information system. In the second part, the improvements of the core data model are raised theoretically and illustrated through a new case study, the archaeological modelling of the Weser (B) drainage basin.

2. The Virtual Leodium project
The initial purpose of Virtual Leodium was the development of an Urban Archaeological information system based on a city scale model. The scale model represents the city of Liège at the beginning of the 18th century on a scale of 1/1200 on around 1m². Its creator Gustave Ruhl, an amateur of archaeology took more than ten years to model this masterpiece of art. The model dates back to the beginning of the 20th century and is an important historic source for the city of Liège before some major changes
occurred in the city due to the industrialization period. The access to the masterpiece of art, accommodated since 1907 by the University of Liège, has for long been restricted due to three reasons: its location (in a manuscripts reading room), its fragility (impossible to move), and its conservation condition (requiring special thermo-hygrometric conditions).

When the project set up, its aims were twofold; creating a 3D model of the scale model to preserve the city model and hence guarantee access for a larger public in order to valorize its documentary and didactic value; and furthermore enriching the city model semantically by iconographic, textual and archaeological sources documenting the scale model and in a broader sense the city of Liège itself (Pfeiffer, 2011).

To do so a workflow composed of three different phases, namely the 3D data acquisition that is subdivided in the 3D data acquisition and the semantic data acquisition, the development of an archaeological data model and finally the implementation of the archaeological information system.

### 2.1. Data acquisition

The model has been scanned in 3D using an optical inspection technique (Billen et al., 2009). This result in a 3D points cloud composed of 650 scans that is cleaned up and merged with the 3D software Geomagic Studio. The data reconstruction is performed manually using the outstanding data surface models and photographs taken from all over the scale model.

Before using the photographs as draping texture within 3D graphics software Maya, the photographs are rectified with the software Photoshop. The 3D modelling phase is still in process, for now about 25% of the entire model is modelled in 3D. The 3D data acquisition workflow is proposed in Fig. 1.

![Figure 1: Workflow of the Virtual Leodium project](image)

In the second sub-phase, the semantic data acquisition is performed. For this phase numerous documents (more than 600) left by Ruhl that includes notes plans, sketches, drawings, and photographs documenting the scale model regarding, i.e. its material, scale, and sources of inspiration, are collected and studied (Pfeiffer, 2011). In general, sources documenting the eighteenth century of Liège’s framework are also studied and in a further step integrated into the spatial database.

![Figure 2: Virtual Leodium information UML model](image)
2.2. The archaeologically-based data model

The archaeological object is placed in the center of our model and is the smallest unit of our system. An historical object is defined by its spatiality, temporality and its functionality. This definition relies on the object identity definition proposed by Peuquet (Peuquet, 1994) for geographical object and re-used for archaeology by Rodier (Galiné et al., 2004; Rodier et al., 2011; Rodier et al., 2012). However, the definition used within our data model differs from the one proposed by Rodier. For Rodier, a change of one of the three constitutional components of an object implies a change of the object identity. We opted for a more flexible identity definition, which means that a change impacting one of the three constitutional components does not necessarily imply an object identity change. This definition relies on the proposition of Hornsby and Egenhofer (Hornsby and Egenhofer, 2000). The second difference concerns the data imperfection that is not handled in Rodier’s approach. We adopt imperfection types of De Runz (De Runz, 2008). First, the presence of the three constitutional components (spatiality, temporality and functionality) is made optional in order to manage the incompleteness of the object. Then, we create a concept called Version, which allows representing multiple spatial representations for the same object and consequently is a way to manage spatial ambiguity. The Modification Event is a concept based on the event management, a quite common concept within the field of computation. With the Modification Event it becomes possible to keep track of the changes of an archaeological object in a large temporal framework. In such way that changes between different states of an archaeological object are recorded.

The model is developed according to the multimodal character of the archaeological data. This means that it is supposed to handle the three most common data sources within archaeology, namely: the written sources, the iconographic sources and the archaeological remains, although for now the implemented prototype of the AIS handles only written and iconographic sources. The Source can document a building, a building element or a group of building but also a Modification Event related to an archaeological object.

The other objects of the model are picked up from the OGC standard CityGML (Gröger et al., 2012) for sake of interoperability with 3D geographical information model.

2.3. Archaeological Information System Prototype

The archaeologically-based data model is mapped into an ArcGIS Geodatabase; it is the core of the Archaeological Information System (AIS) prototype developed during the Virtual Leodium project. The AIS relies on client-server architecture and is developed in Java. The ArcGIS Engine provides the geo-database functionalities. The ArcScene API provides a visualization platform for the 3D data. The previously collected geometric and semantic information is thereby stored in an ArcGIS Geodatabase. Currently only the simplified version of the archaeologically-based data model is implemented. Interested readers are invited to refer to Billen (Billen et al., 2012) to find out more information about the simplified data model.

AIS deals with a two-component interface. On one hand the user can visualize and navigate through the 3D model, on the other hand the user can query the related semantic database. Due to the synchronization of both components, the system provides semantic information about the different buildings by clicking or selecting them, concurrently querying the semantic database highlights the related 3D objects in the 3D scene. The system handles partly the concept of ambiguity and incompleteness by supporting multiple 3D versions of the same object.

3. The Weser Project

The archaeological modelling of the Belgian Weser drainage basin project aims at studying settlements and territorial management of the area in Ancient and Early Medieval time. The research objective is to perform spatio-temporal analysis and predictive modelling of the related archaeological finding sites. This requires recording all archaeological structure evolutions and changes over a long term period and also all related scientific opinions or restitutions. We decide to apply the Virtual Leodium’s data model and improve it in order to resolve remaining critical modelling issues. First of them is the Historical Object identity and discriminant component; how can we determinate what causes a change in an object or, on the contrary, what causes the replacement of an object by another one? The second one concerns both time and function’s imperfection. Finally the understanding of the event management in its computational signification ensures that the changes between versions are managed. This prospect is insufficient for the new case study that needs a huge management of events impacting one (or two or even the three) components of the Historical Object.

3.1. The improved data model

The UML class diagram (Fig. 3) of the model has been designed using ArgoUml software. Just like Virtual Leodium’s model, the core data model links to external models or ontologies for geometry, function and time and to authority bases or data RDF for sources, agent and figures. With the exception of inheritance relationship between episode and version and event, all the relationships of the model are composition relationships.

The following explains the main characteristics of the model. The first improvement relates to Historical Object definition. We suggest to define the Historical Object as follows: a consistent group of elements belonging to the same body from its emergence until its disappearance. The body in question can be an architectural body, a professional corporate body, a human body, etc. In an archaeological sense the body can be seen like a consistent group of facts and units of stratigraphy. This definition is
either close to the “structure” definitions of M. Bats (Bats et al., 1986) and P. Van Ossel (Van Ossel et al., 1988). Indeed, adapting for the Lattes (F) excavations the Harris principles of archaeological stratigraphy (Harris, 1979), Bats defines a structure as a group of contemporaneous facts sharing a same function. Two years later, Van Ossel gives up the contemporaneity to keep only the function’s uniqueness. To us, neither time, space nor function is predominant to discriminate the identity of an object. What matters is the belonging to a same body. According to Emmanuel Kant’s Theory of knowledge (Kant, 1781), the subject constructs its knowledge object, not the contrary. Following this opinion, we think it is the user’s decision and responsibility to establish what type of body he wants to work with in a single project taking into consideration its specific requirements. Besides object identity definition, we wanted also to encompass both function and time lack of fulfillment, incoherency, uncertainty, imprecision, contradiction, like we did for geometry in the Virtual Leodium model. For that reason, the version concept has been generalized to the function component.

The version is a Historical Object state, documented by sources from which we know something about the function and – or the geometry and the chronology of a Historical Object state. Taking into account the fact that chronology is linked with the generalized class episode, we can consider that version is composed by geometry, function and time. But we consider also that each of these components may stay empty because of a lack of documentation. A version exists from the moment that a source tells something about the state of a Historical Object. If different sources describe the same geometry, with the same function and time, it is the same version. If there are differences between geometry, function or time, there are new versions (therefore new episodes) of the Historical Object. Two versions are thus used to express the fact that there are two different functions associated with the same geometry and chronology, or two different geometries associated with the same function and chronology or two different chronologies for a same association function – geometry. A new version will occur for a geometrical information improvement for example.

Compared to the Virtual Leodium model, the event concept is here expanded. We leave the computational vision exclusively dedicated to record changes between two different states of the system and we adopt the cognitive approach where events are used as phenomena descriptor (Pfeiffer et al., 2013). Therefore, an Event is a documented fact, possibly interacting with an HO state but without necessarily generating a change. Nevertheless, it can act as versioning for tracking changes between versions as well. The more interesting property of Event is to manage Historical object’s important data devoid of geometry. Event has neither geometry nor function. It has a more historical nature. It can be seen like the Harris “feature interfaces” units of stratigraphy. However all historical facts recorded for an HO are not necessarily events; most of them are just function’s expression and should be recorded as such.

Episode class generalizes Version and Event classes. It links them not only with Time but with Source as well. It centralizes the whole model around these two most specific classes of historical domains. Although source class must absolutely be filled to allow episode’s existence, time can remain empty. With these characteristics and with the help of version and event, episode ensures management of a proportion of imperfect data, notably the lack of fulfillment, imprecision and incoherency.

Interpretative sequence is an ordered series of episodes. It has one or more sources. The later has one or more
authors, most often historians, archaeologists or architects, all of them involved in studying the historical object story and recording or publishing their interpretations or restorations. However, some sources may be anonymous. Information they provide must be recorded as well. Interpretative sequence has only one direction. It displays a linear or a cyclic time but doesn’t admit flashbacks or contemporaneousness. An interesting aspect of interpretative sequence is that it is the way to deal with relative chronology. It is the way to encompass contradictory information, incoherency or doubt as well: if two people don’t agree about a building history it is not an issue; two different interpretative sequences would be set up.

LifeMap amounts to the Historical Object Timeline. It organizes all its episodes according to a chronological grid despite their involving, or not, into an interpretative sequence. Indeed it is possible to record floating episodes without integrating them into interpretative sequences. LifeMap can be understood like the historical object’s path, or its route built by the amount of versions and events stored in the system. Besides time linked with episode, LifeMap constitutes therefore a second temporal index that facilitates temporal queries.

In the model, people belong to two different categories: historical figures, associated to historical objects episodes and present (or at least not too far away in the past) professional sources’ authors. Two different classes have been created. Agent is the class storing sources and interpretative sequences authors while Figure assembles historical people, implicated in events.

Sources deal with documentation: direct or indirect, digital or printed, textual or figurative, artistic or scientific, sources are raw material for professional studying past. In historical domains, nothing exists without being documented. Sources provide information about episode and interpretative sequences as well. Their rereading and the re-examination of their metadata lead often professionals to suggest new historical interpretations.

3.2. Case study

A specific case study of the Weser project, the evolution of a single building (currently a church): ST-0101 (Fig. 4), will help understanding the use of the proposed model. In this example, the body type selected to define object’s identity is building. Although most of the buildings can be described with names, the identifier of each structure follows the same model: ST (abbreviation for structure) and one number. This choice avoids affecting the data with some dominant interpretation; the case of ST-0101, described below, through light on the importance of semantic neutrality. Nowadays, this cultural heritage building is known under the name “Saint Hermes and Alexander church” or “Theux’s church”. However, it is not that certain this building has always been a Place of Worship. Moreover, it had not always been dedicated to Saint Hermes and Alexander.

Figure 4: Saint Hermes and Alexander Church, North side. ©KIK-IRPA, Bruxelles

Figure 5: General Evolution of ST-0101 building (Bertholet and Hoffummer 1986, p. 256)
In the church’s story state of art proposed by Bertholet and Hoffsummer (Bertholet and Hoffsummer, 1986), the authors give new scientific data (i.e. archaeological excavations or dendrochronological analysis). They used them as a starting point for their new building’s evolution reconstruction divided into nine steps (Fig. 5). They date back to fifth century the very first building of the structure, a house or a pagan Place of Worship (Fig. 5 no. 1). Secondly, during the sixth or seventh century the authors think that the building was extended to become a Christian church, probably dedicated to Saint Pierre (Fig. 5 no. 2). In the second half of the ninth century this building was replaced by a bigger one (Fig. 5 no. 3): a long single-nave building ended by an oriental choir topped by a tour. According to Bertholet and Hoffsummer two events could explain this new building: the receipt of St Hermes’ relics (around 860) and the new church’s dedication, or a fire caused by a Norman’s incursion. Around 1091 a new church, three times bigger than the previous one, was erected and also dedicated to Saint Alexander. It is a three naves hall-church covered by flat ceiling (Fig. 5 no 4). More than one century later in the beginning of the thirteenth century, a tower is added against the north wall (Fig. 5 no 6). From that moment, the church acts as fortified church. During the fourteenth century, important fortification works are achieved: tower is equipped with hoardings (currently still preserved), naves roof are repaired and wooden turrets surmount the angles and the entrance (Fig. 5 no. 6). At the beginning of the sixteenth century, two transformation phases (Fig. 5 no. 7 and 8) modify the east part of the building to give it a gothic shape (Fig. 6). The fortified church use decreases until the eighteenth century phase (Fig. 5 no. 9) when the defensive elements are dismantled.

Using the proposed model, these nine steps scenario can be translated into one life map composed by eleven episodes (nine versions and two events) (Fig. 7). On that figure, light rectangles are versions while dark ones are events. Their first line gives their identifier, the second one their function (for version) or their category (for events), the third line gives time information and the fourth line (for versions only) points to geometrical data.

Four concurrent interpretative sequences organize the episodes series (Fig. 8) given that the first building is assumed to be either a Place of Worship or either a house and given also that two different events may have caused the building of third church.

Moreover, this model allows the management of Interpretative sequences in large numbers. It ensures notably the integration of all scientific opinions related to a same structure. For example, the structure ST -0101 illustrated above has been re-examined on the occasion of artefacts analysis undertook at the European Center of Archaeometry of the University of Liège published in 2009 by Van Wersch (Van Wersch et al., 2009). In her study, the author suggests to revise the dating of two first buildings. Consequently two new versions and four new interpretative sequences appear in the HO Life-Map (Fig. 9).

4. Conclusion and future developments
In this paper, we present the latest improvements of our Archaeological Information System. First, we exposed the general research context of the Virtual Leodium project while insisting on the general workflow and the current prototype. We outlined our current data model by insisting
ont the main concepts, namely the concept of Version that enables partly the handling of data imperfection and the object identity that we based on the main characteristics of the archaeological data. However, some unresolved issues remained in the model, notably the data imperfection management which is not yet completed. Furthermore, the object identity definition still needed to be improved to encompass its vague definition and the event management that was too restricted in regard to the archaeological data complexity.

In respect to the Weser project it turned out that the model had to be improved. In this research, we focused on the object identity definition which we enlarged by introducing the concept of Body. Thanks to this new modeling approach the HO identity integrates henceforth not only the archaeological data characteristics but also the related contextual information in form of scientific interpretations by the use of Interpretative sequence and the concept of Lifemap. Besides we improved the definition of the event management by replacing our computational oriented vision by a phenomena descriptor vision. Due to these improvements and the addition of the Episode concept, a temporal indexing becomes possible.

Applying the model to the case study of the Weser showed that the model manages complex archaeological data. Nevertheless, further case studies need to be performed in order to test the consistency of the model. In the near future, we plan to implement the model. The database will be populated with multimodal data coming from the Weser project. Moreover, we plan to comply at existing modelling standards and to compile our system with Open Source software mainly for interoperability and persistence purposes.
Figure 9: ST-0101 last LifeMap and last new interpretative sequences

Used software:

Adobe System:

ArcGIS Engine:

ArgoUML:

Autodesk:

Geomagic:
Bibliography


