Archeologia e Calcolatori
Supplemento 8, 2016, 254-262

OPENARCHEO2. AN ARCHAEOLOGICAL KNOWLEDGE BASE

1. Introduction

During the ArcheoFOSS 2014 conference I focused mainly on a number of problems which, in my opinion, are currently suffocating the application of digital technologies to archaeological research. In a nutshell, these can be resumed in two statements:

– a long-lasting lack of real methodological innovation: in most cases, the papers given nowadays at conferences are technologically updated, but conceptually not very different from the ones that were presented 15 years ago;

– a scarce contribution to the production of historical knowledge: it still happens too often that the digital solutions, at least in the complete and usually quite elaborated form in which they are presented to the public, remain nice stylistic exercises with a minor real reflection as everyday work tools and, what is even worse, with almost no impact on knowledge production processes.

However, the central subject of this paper is the archaeological information system OpenArcheo2[1]. As a premise it is important to note how the basic idea of OpenArcheo2 is that of representing archaeological knowledge in its broadest sense. This forces us to maintain a constant level of high abstraction in the solution design, implying a description of all primary objects not only as physical features (i.e. the usual approach for archaeological information systems), but mainly in terms of interpretation.

In other words, it means shifting our priority from data to knowledge. In this sense, OpenArcheo2 can be seen as a knowledge-base, a digital solution for storing, sharing and interpreting complex structured and unstructured data (or rather "facts") of a knowledge domain[2].

[1] OpenArcheo2 is currently part of two projects. One is financed on the relevant national interest research program PRIN 2010-2011 and is titled "Global archaeology and history of the rural landscapes of Italy between Late Antiquity and the Middle Ages. Integrated systems of sources, methods and techniques for a sustainable development" (the Sienese Research unit focuses on "Archaeology of Tuscan landscapes between Late Antiquity and the Middle Ages. Research, applications and web 2.0"). The second is carried out in collaboration with the Municipality of Siena and concerns the project "ARCHEOMEDSITES: Safeguard, valorisation and management quality. Use of the management models for the archaeological sites and urban contexts" within the ENPI CBCMED-Mediterranean Sea Basin Programme 2007-2013.

[2] Knowledge-base theory originates from research on AI and expert systems where they have been used, together with inference engines, in decision-making and knowledge production processes (machine-readable knowledge-bases). On the other hand, human-readable knowledge-bases, especially in their semantic-oriented meaning, are the closest ones to our approach. See Kumar 2008, Chapter 4 on knowledge management (especially 113-117).

2. Conceptual model

The conceptual model of our digital solution (Fig. 1) wants to be in first place an exercise of methodology and pragmatic theory (Fronza 2015 for a more in-depth description of the conceptual model and a requirement analysis which explains the main ideas underlying OpenArcheo2). It comes straigth from an everyday practice of archaeological knowledge production processes, as well as from the study of material culture, with a marked topographic approach. Following our research interests, and in order to be as general as possible, the highest level object of OpenArcheo2 has stemmed from the concept of Landscape, understood (in brief) as a portion of land formed in time by man/environment interaction.

Landscapes are represented in their space/time coordinates and variables on the basis of all possibly available material evidences studied through the methods of multidisciplinary archaeological field and laboratory research. In this sense, OpenArcheo2, in its current declination, can be assimilated to some kind of large container of all historical/archaeological information of which we can dispose for a particular landscape, and all the relationships between these data.

In fact, there is no limitation in the origin, definition or size of a Landscape; its extension, as well as its nature, depend directly on the research interests of single scholars and/or projects. So we can have urban landscapes, representing only one city; or landscapes based on historical regions, on an environmental/climatic coherence, on geographic/morphological features, on a particular economic activity, on specific cultural aspects, and so on. It does not even have to be continuous in space (for example we could study the landscape of river fishers in Iron Age Europe and in this case it would be a punctiform collection of territories).

Historical landscapes have been divided into two types of objects:

– Landscape Elements, to be conceptually understood as the high level (interpreted) parts that form a particular Landscape; this means that landscape elements can be, for example, settlements, anthropic exploitations, natural environments, networks of various kinds crossing and marking a territory, etc. Ultimately, these objects form the diachronic plot of the knowledge we have of a specific landscape, which can be more or less dense depending on the data at our disposal and their possible interpretations.

– Topographic Elements, which represent a lower hierarchical level. They are the components that make up each Landscape Element (e.g., buildings, open spaces, productive activities, roads, fortifications, cemeteries, environmental modifications, etc.), of which we have direct knowledge through any kind of historical sources (but, mainly, through material evidences).

Such a classification allows the researcher to master complexity; the informative plans of the space/time context he is studying are represented

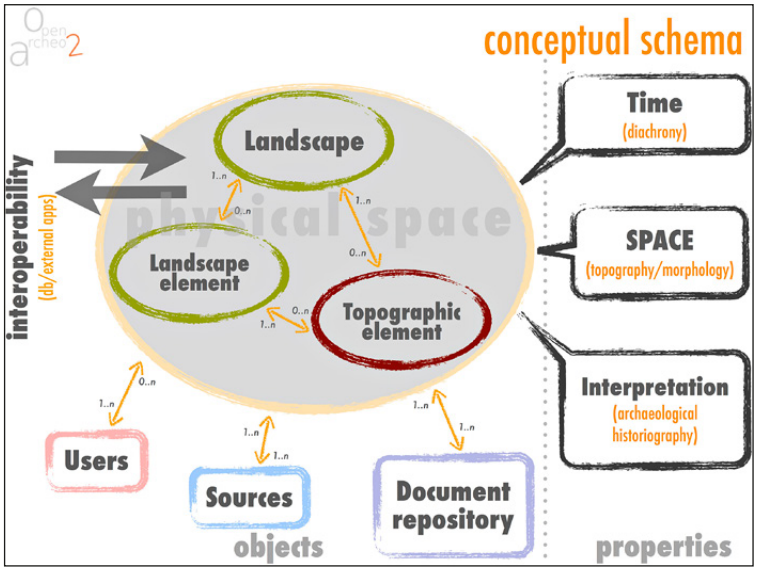

Fig. 1 – The conceptual model of OpenArcheo2.

through their characterizing material components, allowing comparisons between intepretations of different scholars. At the same time, being based on simple concepts, this approach facilitates the use of the platform by a wider and non-specialists audience.

Each one of these three primary objects (Landscape, Landscape Element, Topographic Element) is associated with three groups of basic properties that describe and specify them: time, space and archaeological interpretation. The spatial data is defined as the geographical position and morphology of an object, including its variants or transformations related to the chronological factor and, eventually, to alternative interpretations of the same evidence; an object can therefore have more than one spatial representation. The "time" variable represents the diachronic evolution of an object and is expressed in a numeric form through the concepts of chronological sequences, phases and ranges.

The set of properties we have defined as "archaeological interpretation" derives from the assumption that having a coherent spatial and chronological context does not suffice in itself if we wanto to assign archaeological significance to a material evidence: it must also be "interesting" in the eyes of the archaeologist. In other words, it must implicitly fall into some interpretive grids that reflect the research interests of the scientific community in a certain moment and place. Also, given the highly abstract and social-oriented

approach of the solution, it should be possible to associate multiple meanings to each primary object. This needs to be explained more in detail. In fact, witihin OpenArcheo2, the meanings are handled simply by marking all primary objects with tags and chronotags. In other words, each object will be possibly associated to a set of easy to understand labels/keywords. These tags can be simple or they can have a chronological variable (thus turning into chronotags).

Tags allow the user to place each object within an interpretive grid. The peculiarity of the system, compared to the traditional and large use of tags on content-driven Internet sites, is that these keywords are somehow structured and gathered in tag sets (collections of keywords) which can be created by users. It therefore becomes possible to assign each object to analytical or descriptive tags that are freely conceived by the individual user and relate to a specific domain. Archaeologists, for example, can create their own more or less standardized tag sets reflecting their personal and very specific research interests or that of a scientific community (for example, representing a definition of the settlement type, of a building type/technique, or reporting the presence of artefact classes, etc.).

Such a bottom-up approach encourages the sharing and exchange of ideas and interpretations at various scales (from the single researcher to a small or large research group, to the scientific community as a whole). The ability to create custom sets of keywords configures a totally free system in which researchers can create their own interpretative tools and share them so that others can make them their own. The tag sets in use at any given time will reflect, on the whole, the main interests of the user community; in a sort of natural selection, those most used and shared will survive and continue to replicate, while those more marginal will die with the end of the research interest/project that generated them. The general and wider public also benefits from this tool, having the opportunity to label the historical and archaeological components of a landscape according to the most diverse and personal criteria. Together with the simplicity of the conceptual model and the possibility of commenting any content of the system, this keeps our platform open to any kind of user target, respecting the Public Archaeology principles and the "socially sharable" requirements we decided to adopt.

3. An ontology as a logical model

On a more technical basis, instead of producing a traditional logical model, which would hardly fit our technological and conceptual choice of adopting a strictly object-oriented approach within a NoSQL database management system, we decided to follow a semantic description of the data model. The OpenArcheo2 logical model follows an ontological paradigm, based on objects

Landscape element	*is identified by an*	id	single integer (pk)
	has a	name	single string (100 characters)
	is a	primary definition	single Dictionary lemma (Landscape element.Definition)
	specifically is a	secondary definition	single Dictionary lemma (Landscape element.Definition)
	is located at	main toponym	single Toponym
	belongs to	landscape(s)	collection of Landscapes
	is dating to	chronology	single Chronological range
	is represented by	spatial object(s)	collection of GIS features
	has a	record stamp	single Record stamp
	can refer to	other project(s)	collection of Projects
	has its knowledge built on	source(s)	collection of Sources
	can have	historical toponym(s)	collection of Historical toponyms (historical=true)
	has a	brief description	single string (2000 characters)
	is composed by	topographic elements	collection of Topographic elements
	can be contained by or intersect	spatial repartition(s)	collection of Spatial zones/repartitions
	can include	decontextualised artefact(s)	collection of Decontextualised artefacts
	can be described by	tag(s)	collection of Tags
	can be described by	chronotag(s)	collection of Chronotag refs
	can be further described by	text(s)	collection of Descriptive texts
	can be referenced in	bibliography	collection of Bibliographic titles
	can be represented by	image(s)	collection of Images
	can have attached	repository document(s)	collection of Repo documents
	can have	comment(s)	collection of Comments

Fig. 2 – The "Landscape Element" object class in the OpenArcheo2 ontology.

classified as: primary classes (the main concepts of the information system), secondary classes (conceptually independent, but not primary entities), service classes (used to describe primary and secondary objects). Each object class is

Source (information origin)	*is identified by an*	id	single integer (pk)
	has a	name	single string (100 characters)
	can have an	acronym/abbreviation	single string (10 characters)
	is of	type	single Dictionary lemma (Source.Type)
	is produced by a	main producer	single Producer refs
	is the main subject of an	investigation	single Investigations
	can be the subject of	other investigation(s)	collection of Investigations
	can have a	chronology	single Chronological range
	covers a	geographic area	single Toponym
	covers a	chronological range	single Chronlogical range
	has a	record stamp	single Record stamp
	can refer to	project(s)	*Method: show all projects of related investigations*
	has a	brief description	single string (2000 characters)
	can form the knowledge of	decontextualised artefact(s)	collection of Decontextualised artefacts
	can form the knowledge of	landscape element(s)	collection of Landscape elements
	can form the knowledge of	topographic element(s)	collection of Topographic elements
	can form the knowledge of	toponym(s)	collection of Toponyms
	can be related to	other source(s)	collection of SourceRels
	can be described by	tag(s)	collection of Tags
	can be described by	chronotag(s)	collection of Chronotag refs
	can be further described by	text(s)	collection of Descriptive texts
	can be referenced in	bibliography	collection of Bibliographic titles
	can be represented by	image(s)	collection of Images
	can have attached	repository document(s)	collection of Repo documents
	can have	comment(s)	collection of Comments

Fig. 3 – The "Source" object class in the OpenArcheo2 ontology.

described in a table made of four columns (Fig. 2). The first column contains the domain class (that is, the object class itself which is being described).

The second column represents a predicate linking the domain class to a rank class (property), which occupies the third column. In the fourth column

Bibliographic title.Type of publication	
English	Italian
Volume	Volume
Paper in miscellaneous volume	Articolo in miscellanea
Paper in journal	Articolo in periodico
Multimedia publication	Pubblicazione multimediale

Fig. 4 – OpenArcheo2 simple dictionary example: "Bibliographic title.Type of publication".

there is a logical representation of the rank, describing exactly to which class or simple type (string, integer, etc.) it belongs to; in alternative this column can refer to particular methods which allow us to obtain a specific property. Plural and singular forms of the rank objects are used to indicate the cardinality of the established relationship. For example: "Topographic element is represented by Spatial object(s)" means that each topographic element has one or more spatial objects which represent it (1..n); on the other hand: "Topographic element belongs to a Landscape element" means that each topographic element always belongs to one and only one landscape element (n..1). The 0..n cardinality is expressed through the "can" form of the predicate: thus, "Topographic element can be described by Tag(s)" expresses the 0..n cardinality (topographic element can have from no tags to n tags associated to it), while "Topographic element can have a name" expresses a 0..1 cardinality (topographic element can have only one name or no name). Finally, complex n..n relationships are represented as bidirectional 1..n relationships; for example: "Landscape Element has its knowledge built on Source(s)" and "Source can form the knowledge of Landscape element(s)" (Fig. 3).

A second section of the ontology lists all dictionaries associated to the single rank classes. They are presented in English and in Italian language, following the multilingual approach of OpenArcheo2. But, above all, they can be simple or hierarchical. The former are straight lists of lemmas which compose the vocabulary; for example the dictionary "Bibliographic title.Type of publication" is made of four values (Fig. 4). The latter are based on two hierarchical levels where primary lemmas are further specified by secondary lemmas, corresponding to different properties/rank classes of the domain class which uses the dictionary (a clear example can be seen in Fig. 5, representing the "Landscape element.Definition" dictionary).

The OpenArcheo2 ontology reflects the name and the approach of our solution, plainly aimed at creating a free and open tool for archaeological research. The project and the sourcecode, released under an MIT license, is hosted on Github (http://github.com/scarpazi/oa2/). It is currently being developed mainly in Wakanda (http://www.wakanda.org/), an object-oriented environment. Two different approaches are used for the webGIS component:

Landscape element.Definition			
English		Italian	
Primary definition	**Secondary definition**	**Definizione primaria**	**Definizione secondaria**
Natural environment	Hydrography	Ambiente naturale	Idrografia
	Woods/Fallow land		Bosco/incolto
	Wetlands/Marshlands		Area palustre
	Environmental destruction		Dissesto ambientale
	Geological element		Elemento geologico
Anthropic network	Settlement/demic network	Rete antropica	Rete demico/insediativa
	Economic/productive network		Rete economico/produttiva
	Cultual/religious network		Rete cultuale/religiosa
	Military/fortifications network		Reti militari/fortificatorie
	Communication network		Reti comunicazione
	Infrastructural network		Rete infrastrutturali
Settlement	Simple rural settlement	Insediamento	Insediamento rurale semplice
	Seasonal/temporary rural settlement		Insediamento rurale stagionale temporaneo
	Complex rural settlement		Insediamento rurale complesso
	Urban settlement		Insediamento urbano
	Complex religious place		Luogo di culto complesso
	Necropolis		Area cimiteriale
	Temporary abandonment		Abbandono temporaneo
Anthropic exploitation	Agricultural space	Sfruttamento antropico	Spazio agricolo
	Breeding space		Spazio allevatizio
	Mining area		Area estrattiva
	Collection/supply area		Spazio di raccolta/approvvigionamento
	Hunting/fishing space		Spazio venatorio o di pesca
Functional element	Transformation/processing productive space	Elemento funzionale	Spazio produttivo di trasformazione/lavorazione
	Commercial space		Spazio commerciale
	Simple religious place		Luogo di culto semplice
	Military/fortification element		Elemento militare/fortificatorio
	Hydraulic infrastructure		Infrastruttura idraulica
	Communication infrastructure		Infrastruttura di comunicazione
	Environmental modification		Sistemazione/modificazione ambientale
	Other functional space		Altro spazio funzionale
Planning	Territorial planning	Pianificazione	Territoriale
	Urban planning		Urbanistica
Undetermined	Undetermined	Non determinabile	n.d.
	Generic natural environment		Generico ambiente naturale
	Generic anthropic presence		Generica frequentazione antropica

Fig. 5 – OpenArcheo2 hierarchical dictionary example: “Landscape element.Definition”.

a simple one combining the Wakanda object store and the Leaflet library for online mapping (http://leafletjs.com/) has already been written, while a more complex one, yet to be developed, will be based on Geoserver (http://geoserver.org/) and OpenLayers 3 (http://openlayers.org/).

One last thing to add is that the actually ongoing development phase is a challenge we would gladly share with all interested researchers, even though we cannot help noticing that the constantly increasing use of general purpose OS/FS in archaeology is not matched by a corresponding growth of archaeological software development communities. And that is a shame.

Vittorio Fronza
Department of Historical Sciences and Cultural Heritage
University of Siena
scarpazi@gmail.com

REFERENCES

Fronza V. 2015 (in press), *OpenArcheo2. Da base di dati a base della conoscenza*, in P. Arthur, M.L. Imperiale (eds.), *VII Congresso Nazionale di Archeologia Medievale (Lecce 2015)*, Firenze, All'Insegna del Giglio.

Kumar E. 2008, *Artificial Intelligence*, New Delhi, I.K. International Publishing House.

ABSTRACT

The paper deals with a brief description of OpenArcheo2, a wholly new archaeological information system currently being developed. The system subverts the usual perspective of solutions dedicated to the management/analysis of "raw" archaeological data, focusing entirely on interpreted information. Representation of archaeological knowledge becomes, therefore, the primary objective of the system, as can be clearly seen from the conceptual model and the ontology concisely presented in this paper.