Structured Wiki with Annotation for Knowledge Management: an Application to Cultural Heritage

Eric Leclercq, Marinette Savonnet

Structured Wiki with Annotation for Knowledge Management: an Application to Cultural Heritage

Éric Leclercq and Marinette Savonnet
LE2I Laboratory UMR CNRS 5158 - Université de Bourgogne
9, Avenue Alain Savary - 21078 Dijon, France
Eric.Leclercq@u-bourgogne.fr, Marinette.Savonnet@u-bourgogne.fr

ABSTRACT

In this paper, we highlight how semantic wikis can be relevant solutions for building cooperative data driven applications in domains characterized by a rapid evolution of knowledge. We will point out the semantic capabilities of annotated databases and structured wikis to provide better quality of content, to support complex queries and finally to carry on different type of users. Then we compare database application development with wiki for domains that encompass evolving knowledge.
We detail the architecture of WikiBridge, a semantic wiki, which integrates templates forms and allows complex annotations as well as consistency checking. We describe the archaeological CARE project, and explain the conceptual modeling approach. A specific section is dedicated to ontology design, which is the compulsory foundational knowledge for the application. We finally report related works of the semantic wiki use for archaeological projects.

KEYWORDS

Annotated database, Cultural heritage application, eScience, Ontology engineering, Semantic wiki.

1 INTRODUCTION

Scientists produce more and more data, for example Galperin and Cochrane [1] have counted 1330 databases containing more than two petabytes of data covering different aspects of cell and molecular biology. This data intensive science, called eScience, takes a "data driven" approach, where knowledge emerges from data as opposed to a more traditional "knowledge-driven" approach that examines hypothesized patterns expected from data [2]. To meet the eScience applications we must take into account several characteristics (figure 1):

1) Domain knowledge that can be represented by thesauri, domain ontologies (for example Gene Ontology in biology or CIDOC-CRM in cultural heritage), standards (for example FuGE[1] in functional genomics) and recommendations. Domain ontology is specialized into application ontology. Knowledge is continually changing requiring a flexible data structure;
2) Know-how that is expressed by the business process. Generally, scientific process could be described using five steps: a) data acquisition and modeling, b) collaboration, c) analysis and data mining, d) dissemination and sharing, e) archiving and preservation;
3) Technical basis that often takes the form of collaborative platforms with sophisticated technologies (e.g. the ability to integrate other services such as visualization tools, or spatial analysis tools).

[1]FuGE : Functional Genomics Experiment `http://fuge.sourceforge.net/`

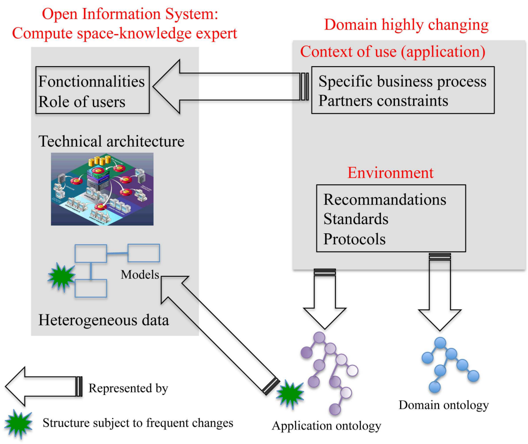

Figure 1. Working environment.

Archaeology is on the verge to embrace eScience [3]: excavations generate exponentially more massive datasets, spurred by the increasing use of imaging tools. Documents (archives, excavation reports) are the basis of the archaeologists work. To meet the needs of archaeologists, we have developed a semantic wiki, *WikiBridge*, where knowledge takes the form of annotated database coupled with a triple-store and a wiki as user interface.

The rest of the paper is organized as follows: section 2 gives an overview of annotation and implementation in two major kinds of systems, section 3 describes the requirements and WikiBridge architecture, section 4 describes our annotation model which is compared to other annotation systems. Section 5 gives an overview of the CARE project and describes semantic tools for archaeology, and section 6 is a related works description. Finally, section 7 concludes the paper.

2 STATE OF ART

Annotations of resources (documents, images, data, web resources, etc.) can be created using different models from simple text to RDF graph. New web based applications such as semantic blogs, social networks, semantic wikis, or data sets providers (for biological, georeferenced, climate data) make an extensive use of annotation systems. Moreover, applications use annotations in different ways [4] for example for finding communities in social network or as a basis for semantic queries in big data applications. Currently no unified model exists for all these kinds and usages of annotations. In this section we first study annotation models and then two kinds of application that make an extensive use of annotations.

2.1 Annotations

Generally speaking, the term annotation refers to a piece of data associated to another piece of data. In software applications the term annotation can denote both the process of annotating a resource or the result of the process [4]. In this article we use annotation for the result of the process.

In document management systems, annotations can be used at different levels: from the whole document to the word level. Annotations can be manual i.e. made by a person, semi-automatic i.e. based upon suggestions or fully automated. Annotations can be associated to a group of users (experts, novice, etc.) and shared with the same group or with other groups.

Annotation creates a relationship between resources denoted by URIs. It established a typed relation between the annotated data and the annotating data. The set of all annotations related to the same resource take the form of a graph structure.

The different models of annotation used in web based applications share a common basis organization in a three-dimensional space: a subject (the annotated data), a predicate (the typed relationship between the annotated data and the annotating data), an object (the annotating data). This conceptual model can be implemented using RDF triples, binary predicates of the first order logic, conceptual graph or semantic network.

Oren et al. in [4] differentiate three types of annotations: informal, formal and ontological. Informal annotations do not use a formal language and thus are not machine-readable. Formal annotations use formal languages that are machine-readable but which do not refer to a common knowledge and thus are not machine-understood. Ontological annotation uses ontology terms that correspond to the conceptualization of a shared knowledge. Therefore, ontology based annotation are machine-readable and machine-understood.

2.2 Annotated Database

In scientific projects, there is a growing need to associate annotations with the corresponding data. Several works related to annotation integration in relational database have been made during the last decade.

The DBNotes system [5] proposes an annotation mechanism for relational databases where each attribute in a relation has a corresponding attribute to hold annotation. DBNotes also, extends the SQL language with a PROPAGATE clause which allows users to specify how to propagate the annotations along with the query answers.

MONDRIAN system [6] introduces an annotation model based on blocks to annotate a set of values. It defines an algebra that allows querying data based on annotations.

bdbms system [7] allows annotations to be defined at multiple levels of granularities (relation, tuple, column and value levels). Moreover, it allows a user relation to have multiple annotation relations attached to it. Each annotation is attached to a region (i.e. defined by a bounding box) that represents the covered area of the annotation. SQL has also been extended to take into account the annotations in the result of a query.

Curated databases are a specific kind of annotated databases [8]. They are expensive to establish, because experts must manually check each data. Data including past versions are recorded as well as data provenance (source of the data) and annotations describing opinions of the experts. The added value of curated databases lies in their quality and organization.

2.3 Semantic Wiki

In traditional wiki, semantics is implicitly described by links between pages and by the context of the link (surrounding text). A semantic wiki is a wiki that makes explicit the semantics, it includes semantic web technologies to enable annotation of resources. Semantic wikis can be built on top of existing wiki or created from scratch. In [9], authors have identified two approaches of semantic wiki: 1) wiki centric approaches use the wiki to organize knowledge i.e. ontology emerges from the wiki through categories and links (*wikis for ontologies*) or 2) ontology based approaches allow importing an existing ontology and using it in the annotation process (*ontologies for wikis*).

In the most popular wiki, MediaWiki, categories are the simplest form of annotation. They are used to classify wiki pages. Semantic MediaWiki [10] is a project that extends MediaWiki and provides new features such as: 1) relations to describe relationships between two pages by assigning annotations to existing links and 2) attributes that allow users to specify relationships between pages and literals. Table 1 gives an example of a page in MediaWiki using links and categories and the same page using Semantic MediaWiki capabilities.

`The city of Moulis is located on the [[Medoc region]]. The building of the High [[Middle Ages]] was discovered in 1993 under the present parish church, largely Romane, surrounded by a parish cemetery until 1901, then transformed in the public square... [Category:ArchaelogicalSite]`
`The city of Moulis is located on the [[region::Medoc]] region. The building of the High [[Middle Ages]] was discovered in [[date::1993]] under the present [[building-type::parish church]], largely Romane, surrounded by a parish cemetery until [[date::1901]], then transformed in the public square... [Category:ArchaelogicalSite]`

Table 1. Concepts, relationships and attributes in MediaWiki and Semantic MediaWiki.

Semantic MediaWiki facilitates entry of data by using the Semantic Forms extension[2]. Moreover, Semantic MediaWiki engine allows to load ontologies and to consult them as wiki pages. OntoWiki [11] has been developed with the main objective of facilitating the acquisition and presentation of data. It offers forms and includes the RDF triples directly in the text using an appropriate syntax. AceWiki [12] follows a different knowledge acquisition strategy using a controlled language ACE (*Attempto Controlled English*). Sentences in ACE are automatically translated into OWL and/or SWRL. AceWiki integrates the OWL reasoner Pellet and ensures that the ontology is always consistent.

IkeWiki and KnowWE are two examples of the second category. IkeWiki [13] is a tool for collaborative knowledge management that requires a pre-existing ontology. The knowledge base is stored using the Jena RDF framework, and a SPARQL engine allows querying it. Two editors are available: one for metadata with a self-completion mechanism and a WYSIWYG editor for content. In addition IkeWiki supports importation of existing content from Wikipedia. Similarly, KnowWE [14], built on top of JSPWiki, uses the Sesame RDF storage.

SweetWiki [9] allows users to tag pages, (called social tagging) and also integrates external ontologies. The set of users tags generates a folksonomy. In addition, SweetWiki adds a WYSIWYG editor for managing content and meta-data, a reasoning engine used for querying the wiki content.

Some semantic wikis have been developed for specific domains, for example, BOWiki for biomedicine and

[2] An extension for MediaWiki `http://www.mediawiki.org/wiki/Extension:Semantic_Forms`

SWiM for mathematics. BOWiki [15] allows to access to several ontologies like the Gene Ontology and ontologies about cell types or anatomy. SWiM [16] extends IkeWiki adding support for the OpenMath language, an XML language for expressing the logical structure of mathematical formula.

In short, the first category of semantic wikis can be used to present knowledge by structuring concepts through pages, categories and links. The second category of semantic wikis based on pre-existing ontologies can be used as a platform to build applications that require a global consensus over knowledge in order to maintain the quality of data.

2.4 Comparison

The annotation tools and structured content provided by template forms [17] found in semantic wikis place them between conventional wikis and databases. A semantic wiki is more structured than a wiki, but at the same time the structure is dynamic and extensible. In a database centric approach, the database schema is built upon entities identified in the first step of analysis, and thus based on an instant knowledge. In domains characterized by a rapid evolution of knowledge, such as biology or archaeology, a static database schema is not suitable and can be proscribed by the cost of evolution. The process of generating the structure is also different. In database the domain modeling is usually done by some experts; in semantic wiki this is a collaborative, dynamic and evolutionary process.
Compared to a traditional database, a semantic wiki allows: 1) to expand the structure of documents content; 2) to enable a data model emergence from the usage; and 3) to support collaborative, distributed workflows and processes (figure 2).

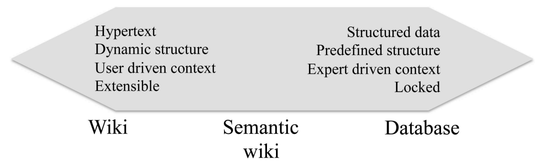

Figure 2. Semantic wikis between wikis and databases.

3 ARCHITECTURE OF WIKIBRIDGE

In a survey authored by Uren et al. [18], authors study semantic annotations, identify a number of requirements, and review some semantic annotation systems. WikiBridge's design principles are following the seven requirements given by Uren et al.: easy to use interface, user collaborative design, support of different user skills, support of heterogeneous format, compatibility with Semantic Web standards, annotation capabilities and storage, and support for reasoning. In the next subsection we develop the most important requirements with regard to the architectural design of WikiBridge.

3.1 Requirements

In a knowledge engineering process, it is common that non-technical domain experts work together with experienced knowledge engineers. To support different levels of user skills certain advanced functionalities should be hidden from novice users but made available to experienced users. Thus, we use an Access Control List (ACL) mechanism to describe privilege control

depending on user identity and group affiliation. Advanced users can define forms to help users to structure wiki articles; each part of a form generates automatically annotations. During the annotation process, a wizard suggests terms of the ontology according to highlighted section in the form.
To be able to exchange data with other applications (e.g. ontology editors, Web services, other wikis), compliance to Semantic Web standards is required. WikiBridge is purely based on existing Semantic Web standards such as the Web Ontology Language OWL for describing ontologies and W3C's RDF for annotations.
We consider reasoning as one of the most important functionalities as it allows: 1) to emerge knowledge that is not explicit in the data; 2) to check the meaning of annotations with regards to the context of the annotation; and 3) to enhance navigation and search.

3.2 Architecture

One of the most famous semantic wiki is Semantic MediaWiki (SMW), which is based on MediaWiki [19]. In 2009, when we have started the project, complex annotation and consistency checking were identified as mandatory functionalities. In 2009, SMW doesn't provide complex annotation and doesn't have consistency checking in its roadmap.
We have started by extending MediaWiki with the following semantic components: form based acquisition interface with automatic annotation, annotations wizard, annotations validation based on the context of a document, semantic rules and a query engine.

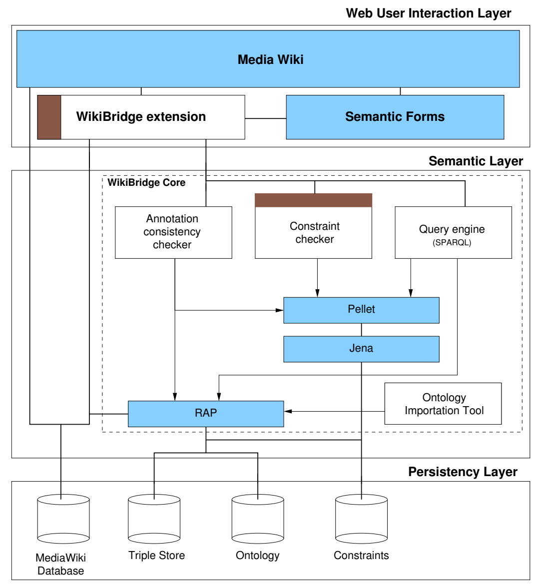

Figure 3. WikiBridge's architecture.

Figure 3 presents the architecture of WikiBridge, the semantic components are structured in three layers, third party components are in blue boxes, and brown boxes include Web services for managing interactions between inner services or for adding query capabilities for external applications.

The **user interaction layer** is covered by MediaWiki and structured data control needed for inexperienced users is managed by Semantic Forms. This extension provides advanced users with a specific description language that allows to define new forms (figure 4). Modules corresponding to the interaction layer are represented on the top of figure 3.
The **semantic layer** manages annotations and query processing. To improve the quality of information during the input and annotation processes, we propose three semantic components in WikiBridge Core (white boxes in figure 3) developed on the top of third party tools (RAP - RDF API for PHP, Pellet and Jena).

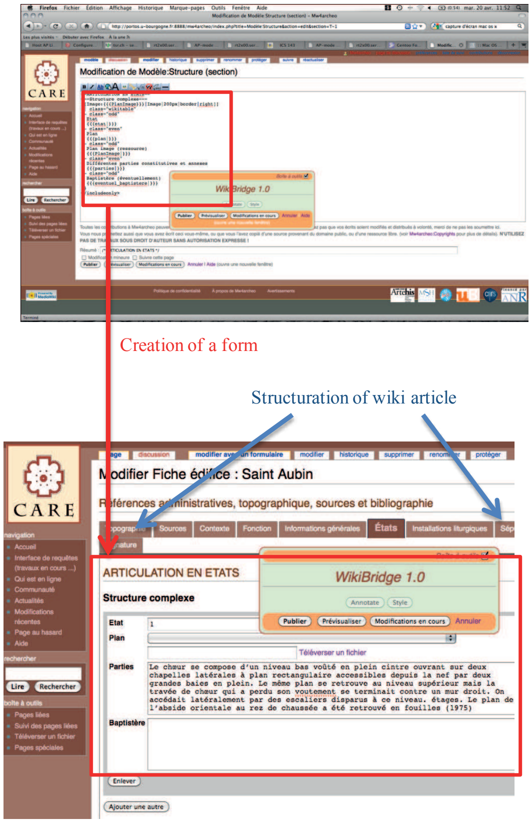

Figure 4. User interaction layer

The syntax and the semantics of annotations, made by experts, are guaranteed by application ontology. An annotation wizard helps users to construct simple or complex annotations by selecting ontology terms in lists and giving them properties and values (figure 5). Ontology terms are retrieved from ontology concepts through SPARQL queries. Annotation construction is a context sensitive process, initial terms displayed to users are directly connected to fields of forms.
Each document is identified by its URL within the wiki and annotations use this URL as a basement for identifying fragments of content.
Simple annotation allows annotating a subject by describing its property using a literal or a reference to an ontology term.
Complex annotation allows annotating a subject with two or more simple annotations and references to other elements (subjects). For example we can annotate an altar with its dimension, its building material, its location in the nave. The nave is detailed in another part of the document.

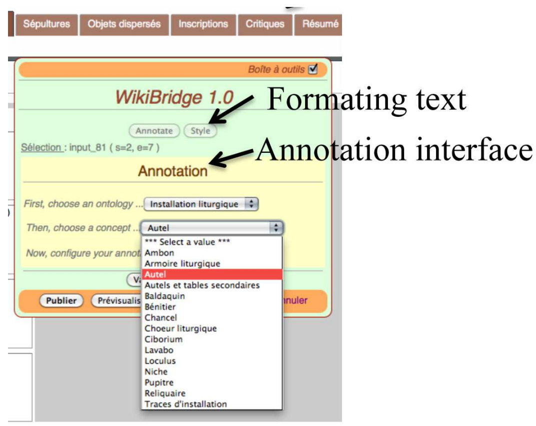

Figure 5. Annotation wizard.

Annotation consistency checking is operated by a set of specific Java components interacting with RAP, Pellet and Jena. Only a subset of first order logic constraints is checked. WikiBridge extension (top of figure 3) connects to the Java constraint checker by the means of a Web service. Moreover, rules can be added to query ontology and annotations in order to test new facts and thus to produce new knowledge that can be inserted in the set of semantic constraints. Two kinds of constraints can be checked by using the ontology knowledge: 1) domain values of properties using *ABox* capabilities; and 2) structural consistency of properties using *TBox* capabilities (for instance, a cathedral can have a nave but cannot have an atrium). Nevertheless, some domain dependent constraint cannot be embedded in the structure. For example "In France, there is no church with rammed earth wall for the studied period however, this technique is used in Ireland and in other countries" can be translated by the following constraint that must remains consistent:

hasRammedWall(?x)∧
edificeType(?x,?t)∧
edificeCountry(?x,?c)∧
c='France'∧ t ≠'church'

The **persistency layer** includes four types of storage (bottom of figure 3): documents content, semantic annotations, ontology, and constraints.
The content of documents is stored by MediaWiki specific database.
Annotations are stored as triple in RAP triple store and they can be retrieved by WikiBridge user layer to display annotations with icons and colors in a document or by the SPARQL query engine.
The ontology imported as an OWL file is stored in a specific schema managed by RAP. Ontology terms can be then queried using SPARQL and results can generate wiki pages.
Constraints are stored in plain text using Jena rules syntax. A type attribute specifies if the rule can be applied to check the ontology structure or to check annotation consistency.

Information access has been designed with taking into account some features about users. We have thus identified a usage typology in accordance to 1) kind of usage (reader, investigator, and annotator); 2) knowledge degree of the domain (domain specialists like historian researchers and non-specialists).

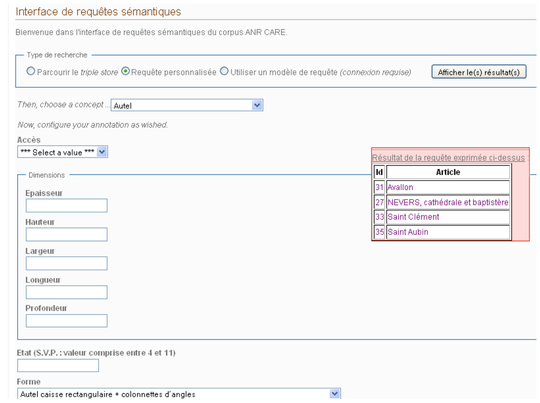

Figure 6. Query interface.

To handle these different types of users, we offer three types of queries:

- Faceted browsing allows users to explore by filtering available information with the ontology structure (figure 6);
- Form based searching provides semantic search by filling in parameters of parametric queries identified during the analysis of requirements;
- Aggregate view for each article, all annotations related to the article are displayed in a factbox.

Nevertheless all types of queries rely on the SPARQL query engine that also allows to process in line queries into wiki pages in order to summarize information.
To operate spatial and temporal analysis on annotations a set of Web services has been developed. Some specific services allow retrieving objects and their coordinates according to a set of conjunctive properties (figure 7). Moreover, a generic Web service has been developed to handle SPARQL queries.

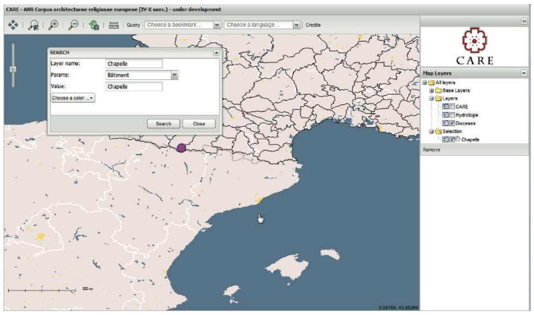

Figure 7. OpenLayers interface interacting with a Web service of WikiBridge.

4 ANNOTATION MODEL

In this section we describe the annotation model used in WikiBridge and then we compare WikiBridge capabilities with other annotation-based system according to Oren's criteria.

4.1 WikiBridge annotation model

Three kinds of annotation can be set in WikiBridge: automatic annotations are provided by template form, assisted annotations are set by users using the annotation wizard and fully manual annotation can be written by experienced users directly in documents using the wiki syntax. These three kinds of annotation share the same basic construct:

$$A=(s,p,o)$$

where s is the subject (the annotated data), p is the predicate (the type of the link) and o is the object (the annotating data). The types of each component are the following:

- s is a URI/URL that refers to the document (i.e. an article in the wiki or a part of an article);
- p is a URI that refers to an ontology concept or property;
- o is a literal or a URI that refers a individual in the ontology or a URI outside the scope of the wiki, or null.

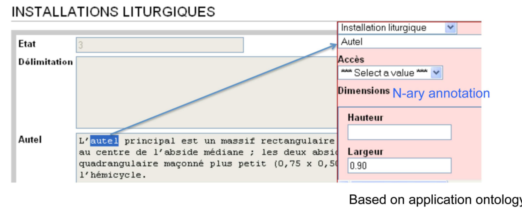

Figure 8. Construction of recursive, n-ary annotation with the annotation wizard.

Our model of annotation allows defining three basic structures of annotation: simple, complex, and recursive.

- A simple annotation has the type *(s,p,o)*. *s and p* cannot be null. If o is null and p refers to a concept, the annotation specifies the type of the subject. It can be viewed as a constraint that is a restriction of an attribute domain. If o is not null it must refer to a literal or to an individual that belongs to the concepts specified by p. It can be viewed as a database constraint that checks that an attribute value is in an enumerate list of values;
- A complex annotation or n-ary (noted `A-cplx` in table 2) is a list of simple annotations related to the same subject (figure 8). All the predicates used in the list must be different;
- A recursive annotation (noted `A-rec` in table 2) is an annotation based on a previous one, used to give details on the object. A recursive annotation has different levels. An annotation of the level i explains the object o of the parent annotation (i.e. from the level i-1). If all the annotations of level i are in list of annotations then all annotations in the list share the same subject (o). The recursive form of annotation is based on the semantic value model defined by Sciore and Rosenthal in [20].

According to the model, the abstract syntax depicted in table 2 is used in the wiki annotation syntax.

```
<A>:=<A-simple>
      |<A-cplx>
      |<A-recursive>
<A-simple>:=(<s>,<p>,<o>)
<A-cplx>:=(<A-simple>,<A-cplx>)
<A-rec>:=((<s>,<p>,<o>)<A>)
<s>:= URI | URL
<p>:= ontology concept
      |ontology property
<o>:= ontology individual
      | literal
      | URI
      | URL
      | null
```

Table 2. A simplified version of the abstract syntax used for WikiBridge's annotations.

An annotation is defined with regard of a context. The context has two dimensions: the common knowledge dimension (i.e. the structure of the ontology and the rules over the concepts), the local knowledge dimension (i.e. the set of all the annotations and sub-annotation which share the same root subject). An annotation is valid only if it is consistent in the two contexts.

Annotation consistency in the common knowledge dimension is checked by evaluating ontology's constraints instanciated by annotations (an example is given on page 8). The annotation is inconsistent if it violates one constraint that uses one of the predicate or object of the annotation.

Annotation consistency in the local knowledge dimension is checked by evaluating intrinsic annotation model constraints. For example a subject cannot be associated to two different objects using the same predicate.

4.2 Evaluation Criteria

Oren et al. in [4] have combined several criteria from the literature as well as their own criteria to classify precisely annotation systems using six properties:

1. Association refers to the way the annotation is associated with the annotated resource (embedded or externally stored);
2. Subject granularity indicates the scope of the annotation (a whole document, a section, a sentence, a word, a numerical data etc.);
3. Representation distinction indicates whether the annotation can distinguished object or concepts in document using a reference system from values;
4. Terminology reuse is related to the level of interoperability of annotation and indicates whether the annotations are ad-hoc or use terms from ontologies;
5. Object type indicates the type of the annotation (literal, textual, structured, ontological);
6. Context indicates meta-data for an annotation (when it was made, by whom, provenance, etc.).

System Property	Annotated DBs	Semantic Wikis	WikiBridge
Association	embedded	current page embedded	current page embedded and triple store
Granularity	relation, tuple, attribute and value	document and fragment	document and fragment
Representation	yes/no	yes/no	yes
Terminology reuse	yes/no	yes/no	yes
Object type	literal	literal, ontology term, URI	literal, ontology term, URI
Context	no/yes	no/yes	yes

Table 3. Comparison of WikiBridge with well established annotation systems.

A comparison between annotated databases, semantic wikis and WikiBridge is shown in table 3. It shows that WikiBridge provides a better representation of semantics with levels of granularity and context representation. Furthermore, WikiBridge supports strictly typed annotation through association, representation and object type that enable consistency checking.

5 PROJECT OVERVIEW

The aim of the international project CARE (*Corpus Architecturae Religiosae Europeae*) is the setting up of a corpus describing Christian edifices in Europe. Italy, Spain, Czech Republic, Poland, Slovakia, France and Croatia have been included in the project four years ago. Each edifice is described in a document that focuses on the definition of states of evolutions from the 4th century to the 11th century.
The French corpus focuses on the 7th and 8th centuries with very rich decades in terms of number of monuments (`http://care.u-bourgogne.fr`).
Archaeology is the "science of destruction", in which the process of excavation removes the sedimentary and cultural context of artifacts and architecture definitively. Thus, the accurate recording of contexts and artifacts are crucial. Representing and managing knowledge in cultural heritage require a deep understanding of specific concepts. Building collaborative platform brings out some challenging characteristics: 1) complexity of data (heterogeneous, incomplete, uncertain, inconsistent, spatial, and temporal); 2) domain knowledge barrier; 3) evolving knowledge; and 4) skills of actors.

5.1 The CARE Community

From an organizational perspective, the CARE project takes the form of an expert network collecting and providing information on edifices, analyzing historical sources, filling documents and collaborating in the exploitation of the corpus through smaller research groups. Additionally, the project involves undergraduate students that help in collecting information but lack expertise required to interpret complex data.
In France, more than sixty researchers are collecting and analyzing data concerning approximately 2700 monuments.
Two key characteristics outline the CARE community:

- Multi-disciplinarity: The data collecting process involves archaeologists, historians, art historians, topographers, draftsmen. It is designed as a collaborative process which merges information from various disciplines;
- Inter-disciplinarity: The interpretation of data brings together all the actors which also enrich their respective practices by the confrontation of methods or problems.

5.2 Conceptual Modeling for the Foundational Knowledge

Linster in [21] shows that the interaction among domain experts, knowledge engineers and tools creates knowledge. In addition, he has shown that the process of elaborating a knowledge-based system is a constructive model-building process that includes: a discussion process between knowledge engineers and domain experts as well as the construction of a conceptual model (i.e. a general and abstract framework). Thus, the knowledge engineering activity encompasses the design of two kinds of models: model to make sense and model to implement systems.

We have applied Linster's guidelines to the CARE project in order to initiate a foundational knowledge from the corpus of documents. The first stage is identification of salient concepts. The key concept is the edifice to which it is essential to model changes. All constituent elements of a building need to be described. They can delimit space or define religious function (baptismal, funerary, etc.). All changes of space or religious function determine a new edifice state. Edifices and their evolutions are described in a set of documents. Salient concepts let us to build a conceptual model in which three groups of elements have been identified (figure 9):

1. Spatial concepts without temporal relationship (light grey): concept EGS refers to edifice, group of edifices or space inside edifice such as nave or apse. Composition relationships can be identified between EGS;
2. Spatio-temporal concepts (grey) called SEGS, represent variations of spatial concepts in time. Spatio-temporal concepts are linked to a date or a period. Dating elements can be determined by documents, or described by methods such as C14, thermoluminescence or stratigraphy;
3. Identification of vocabulary terms for the project domain (black). While concepts EGS and SEGS are used to structure the descriptions, terms are used to describe specific elements including, properties of edifices, religious functions or manufacturing techniques.

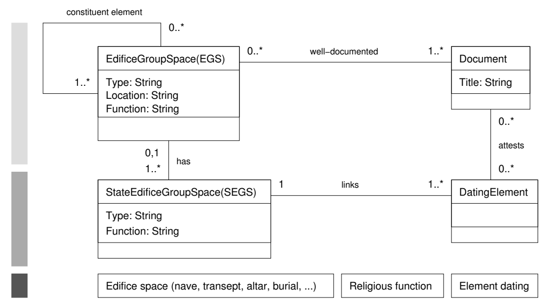

Figure 9. Conceptual model of the CARE corpus.

The description of the design of the ontology is detailed in the next section.

5.3 Offering Semantic Tools For Archaeology: The CARE Ontology

Cultural heritage collections can be annotated with different thesauri. About twenty thesauri are described at the page `http://tinyurl.com/5u8bjer`.
The CIDOC Conceptual Reference Model (CRM) provides an extensible ontology for concepts and relationships in cultural heritage domain (`http://www.cidoc-crm.org`). Since 2006, it is an ISO standard (21127:2006) for exchange of cultural heritage information. CIDOC-CRM aims at treating all types of material collected and/or displayed by museums: sites, monuments as well as collections of fine and applied arts. It is intended to encompass the detailed description both of individual materials as well as groups of materials as a whole. It also covers contextual information: historical, geographical context in which materials are placed and which gives them much of their significance and value. The event notion in CIDOC-CRM is represented by the concept of *event*. An

event describes environment of the material over the time and what could happen to it. The central notion is complemented by: 1) the *TimeSpan* concept describes the moment it happened; 2) the concept of *Place*; 3) who did it (*Actor*); and 4) what is being described. CRM offers notions of *Physical Objects* that can be natural or have been manufactured by man, and *Conceptual Objects*.

Since CRM is a reference in the field, we use it as a starting point to establish CARE ontology. The view of the CARE ontology as a specialization of CRM allows us to relate to a standard and comply with it. The archaeologist uses two sources of information. First, there are material data like pottery, clothing and architectural items like walls, columns, floors. Second, there are descriptive data that help to associate material data found in a location with their context. Descriptive data can be measurements (height, width and length but also mass, density), direction, and association. Association seeks the position of the material data in relation to its surroundings. Accordingly, the CARE ontology has four parts: 1) religious concepts, 2) their spatial relationships and 3) characteristics and 4) timeline to track evolutions (figure 10). Each sub-tree from `THING` may be considered as an ontology of a particular domain. We call a branch of ontology a hierarchy `is_a` with one root. To cover a wider field of knowledge it should compulsorily consider the relationships between several sub-trees. Grenon et al. [22] propose the definition of three kinds of relationships:

- Intra-ontology: relationship with two concepts of same part of an ontology;
- Trans-ontology: relationship with a concept of a sub-tree and a concept of another sub-tree. For instance, a building is consecrated to a saint, in DL we can write:

`Building ⊑∃ isConsecrated.Saint`

- Meta-ontology: relationship with a concept of an ontology and another ontology (considered as a whole).

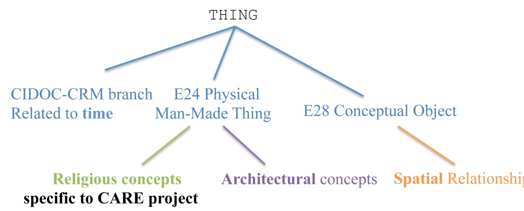

Figure 10. Branches of CARE ontology (in blue CIDOC-CRM ontology, concepts with EXX prefix are CIDOC-CRM concepts).

Modeling religious concepts

Religious concepts in CARE are edifices, with its decomposition into different constituent elements (nave, transept, apse, etc.), liturgical installations (altar, ambo, ciborium, etc.) and burials. These concepts have been placed under the concept `E24 Physical Man-Made Thing` CIDOC-CRM. Indeed, CIDOC-CRM defines this concept as "*all persistent physical items that are purposely created by human activity*".

Modeling characteristics of an edifice

To detail parts of an edifice, we introduced the concept of architectural elements. It describes masonry, floor, opening, inscriptions, as well as construction techniques, dimensions, and colors, etc.

Modeling spatial relationships in archaeology

The geometry implementation in the textual descriptions that are analyzed is a complex geometry. Indeed, these descriptions do not refer to an absolute and orthonormal space: it is rather, a space perception or a cognitive space whose structure is largely based on the functional aspects and objects described, and the perspective of the archaeologist. From the analysis of textual descriptions of religious concepts (description of the position and shape) given by archaeologists we have found four types of spatial properties:

1. Orientation properties: forward, back, bottom, next to, on one side, under, below, at a lower level, at the same level, above, right, left, center, prior, and the cardinal directions;
2. Boundary properties: outside, inside;
3. Distance properties: near, far, next to, around;
4. Topological properties: flank, join, open on the side, link to, stand against, surround, isolated. To represent the topological properties, we used the work of Hegenhofer and Herring [23]. The authors have defined a minimum set of eight relations (disconnected, externally connected, partially overlap, equal, tangential proper part, non-tangential part, tangential proper part inverse, non-tangential part inverse) describing the relations between two regions. This set serves as a basis for defining other topological relation by using a composition.

Modeling temporal relationships to track evolutions

When writing of his excavation report, the archaeologist graphically summarizes the results obtained with a timeline that is often organized by anterior/posterior relationships: materials are considered in relation to each other. In the CARE project, time model is based on following criteria: some absolute benchmarks and a relative chronology based on intervals. We have established a convention for century division and boundaries. Centuries start at year 1 and end at year 100. We also have established subdivisions terms such as early (1→32), mid (33→66) and late (67→100). These century divisions are placed under *E52 TimeSpan*.

CIDOC-CRM offers specific concepts relate to time [24, 25]. Some Allen's relationships [26] are properties used.

The CARE project aims to follow the evolutions (creation, modification, deletion) of an edifice and its constituent elements. The concept of activity is important for CARE project because a state may be characterized by an activity in the CIDOC-CRM ontology. The concept *E7 Activity* is defined as follows: "*The action or sequence of actions intentionally carried out by Actors that result in changes of state in the cultural, social, material systems which interest*". This notion includes both complex and long-lasting actions such as building an edifice, as well as simple and short-lived actions. Following the concepts of the CIDOC-CRM ontology, we use seven concepts, specialized concepts of E7 activity, to model states of edifices: *E6 Destruction, E11 Modification, E81 Transformation, E63 Beginning of Existence, E64 End of Existence, E79 Part Addition and E80 Part Removal.*

The CARE ontology has been designed using Protégé and it actually encompasses 124 classes and 715 individuals.

6 RELATED WORKS

Several semantic wikis have been developed or used specifically for cultural heritage applications.

Witte et al. [27] present an approach to cultural heritage data management which integrates different technologies: a wiki user interface, text mining support using a Natural Language Processing (NLP) framework and ontologies based on OWL and RDF. Authors have implemented the ideas for the *German Handbuch der Architektur*, a comprehensive multi-volume encyclopedia of architecture. A volume (506 pages) of the encyclopedia is converted into wiki pages. Authors have to capture two sub-domains by ontologies: the domain of document management (i.e. sentence, noun, page number, etc.) and architectural domain (i.e. wall, building material, etc.). NLP allows connecting architectural concepts with document-specific one, e.g. sentences that mention construction elements of a certain material. A public version is available at `http://www.semanticsoftware.info/durm`.

The HermesWiki [28] is a semantic wiki in the historical domain in German language. The main objective is to provide an overview on Ancient Greek History for teaching purposes of undergraduate students. The wiki consists of three parts: a collection about twenty essays giving a comprehensive domain walk-through, translations of the describing ancient sources and a glossary. The entries in the glossary are tagged. It has been implemented as a plugin for KnowWE, reusing as much of the core components as possible. A public version is available at `http://hermeswiki.informatik.uni-wuerzburg.de`.

NavEditOW is a framework for ontology driven web site. It has been exploited to support a semantic description of two projects: 1) a web portal and a set of advanced services supporting the sharing of knowledge about Prehistory and Protohistory in the Italian context [29]. In particular, one of the services is represented by a digital library, in which entries (i.e. bibliographic description of publications) will be ontologically described. The system is currently on-line at `http/www.archeoserver.it` and; 2) SilkRoDE (Silk Roads in the Digital Area) project that aims to collect, structure and diffuse all knowledge about the Cultural Heritage of Central Asia from fields such as archaeology, geography or history [30]. The ontological approach provides the required expressiveness and flexibility to support rich forms of navigation among stored contents. The framework integrates a wiki engine for rendering documents stored in the ontological tier.
In the same view, MANTIC is a web application that realizes a portal for archaeological information about the city of Milan [31]. MANTIC integrates different data sources and the global schema is based on CIDOC-CRM.

Our approach of semantic wiki is directed towards scientific application domains, which contribute to produce knowledge [32]. These kinds of applications rely on core ontologies that act as a consensus. Querying and analyzing data enhance knowledge, new concepts can emerge and new constraints can be found out. As a result, ontologies can be modified dynamically and semantic checks are necessary to find inconsistent annotations with regards to ontology.

7 CONCLUSION

Wiki solutions meet the requirements of a web platform with collaborative capabilities. Easy setup and rich editing support are primary reasons for the widespread adoption of wikis. Users can enter text and others types of data (pictures, video) and connect content through hyperlinks. Most of wikis also provides a versioning system to track content changes and a full-text search engine for querying wiki pages. The narrative structure is one advantage of wiki documents centric approach, compared to a database centric approach. Nevertheless, a mere document management system is not sufficient to catch interdependent structures of knowledge. Adding semantic annotation capabilities to documents allows different levels of interpretation and can sustain: 1) knowledge evolution by keeping track of the successive annotations; 2) better quality in the query evaluation process; and 3) amenable result displayed according to user skills. Annotations can be defined at a coarse-grained level or at a fine-grained level. Ontology must be associated to the annotation system to provide a semantics for annotation terms according to domain knowledge. Semantic wiki solutions meet the requirements of annotation system and knowledge description. Adding semantics yields two dimensions of enhancements to a wiki: 1) adding a more formal structure to the wiki; 2) exporting, integrating and reusing information by the adoption of standard semantic technologies. WikiBridge thus seems to combine the best from two worlds: structure from databases as well as expandability and collaboration capabilities from wiki systems.

ACKNOWLEDGMENTS

This research is supported by the French ANR funding (ANR-07-CORP-011).

8 REFERENCES

1. Galperin, M., Cochrane, G.: The 2011 Nucleic Acids Research Database Issue and Online Database Collection. In Nucleic Acids Research, vol.39 (Database issue), pp.1--6, (2011).
2. Newman, H., Ellisman, M., Orcutt, J.: Data-intensive e-science frontier research. In Communications of the ACM, Vol.46(11), pp. 68--77, (2003).
3. Petrovic, V., Gidding, A., Wypych, T., Kuester, F., DeFanti, T., Levy, T.: Dealing with Archaeology's Data Avalanche, In Computer, Vol.44(7), pp. 56--60, IEEE, (2011).
4. Oren, E., Möller, K., Scerri, S., Handschuh, S., Sintek, M.: What are Semantic Annotations?, Technical report, DERI Galway, (2006).
5. Chiticariu, L., Tan, W., Vijayvargiya, G.: DBNotes: A Post-It System for Relational Databases based on Provenance. In Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 942--944, (2005).
6. Geerts, F., Kementsietsidis, A., Milano, D.: MONDRIAN: Annotating and Querying Databases through Colors and Blocks. In Proceedings of the 22nd International Conference on Data Engineering (ICDE), pp. 82--91, (2006).
7. Eltabakh, M., Aref, W. ,Elmagarmid, A., Ouzzani, M., Silva, Y. : Supporting Annotations on Relations. In 12th International Conference on Extending Database Technology (EDBT), pp. 379--390, (2009).
8. Buneman, P., Cheney, J., Tan, W., Vansummeren, S.: Curated databases. In PODS, pp. 1--12, (2008).
9. Michel Buffa, M., Gandon, F., Erétéo, G., Sander, P., Faron, C.: SweetWiki: A semantic wiki. In Journal of Web Semantics, vol.6(1), pp. 84--97, (2008).
10. Krötzsch, M., Vrandecic, D., Völkel, M.: Semantic MediaWiki. In 5th International Semantic Web Conference, pp. 935--942, (2006).
11. Auer , S., Dietzold, S., Riechert, T. : OntoWiki - A Tool for Social, Semantic Collaboration. In 5th International Semantic Web Conference, pp. 736--749, (2006).

12. Kuhn, T.: How Controlled English can Improve Semantic Wikis. In 4th Semantic Wiki Workshop (SemWiki), (2009).
13. Schaffert, S.: IkeWiki: A Semantic Wiki for Collaborative Knowledge Management. In 15th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE), pp. 388--396, (2006).
14. Reutelshöfer, J., Haupt, F., Lemmerich, F., Baumeister, J.: An Extensible Semantic Wiki Architecture. In 4th Semantic Wiki Workshop (SemWiki), (2009).
15. Hoehndorf, R., Bacher, J., Backhaus, M., Gregorio, G., Loebe, F., Prüfer, K., Uciteli, A., Visagie, J., Herre, H., Kelso, J.: BOWiki: an ontology-based wiki for annotation of data and integration of knowledge in biology. In Proceedings of BMC Bioinformatics, Vol.10(s-5), (2009).
16. Lange, C.: SWiM - A Semantic Wiki for Mathematical Knowledge Management. In 5th European Semantic Web Conference, (ESWC), pp.832--837, (2008).
17. Haake, A., Lukosch, S., Schümmer, T.: Wiki-templates : adding structure support to wikis on demand. In International Symposium on Wikis, pp. 41--51, (2005).
18. Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera M., Motta E.,Ciravegna, F.: Semantic Annotation for Knowledge Management: Requirements and a Survey of the State of the Art. In Web Semantics: Science, Services and Agents on the World Wide Web, vol. 4(1), pp. 14--28, (2006).
19. Völkel, M., Krötzsch, M., Vrandecic, D., Haller, H., Studer, T.: Semantic wikipedia. In Proceedings of the 15th international conference on World Wide Web, pp. 585--594, (2006).
20. Sciore, E., Siegel, M., Rosenthal, A.: Using semantic values to facilitate interoperability among heterogeneous information systems. In ACM Trans. Database Syst., vol.19(2),pp. 254--290, (1994).
21. Linster, M.: Linking Modeling to Make Sense and Modeling to Implement Systems in an Operational Modeling Environment. In EKAW, pp. 55--74, (1992).
22. Grenon, P., Smith, B.: SNAP and SPAN: Upper level ontology and framework for formalization. In Spatial cognition and computation, vol.4(1), (2004).
23. Hegenhofer, M., Herring, J.: Categorizing Binary Topological Relations Between Regions, Lines and Points in Geographic Databases. Technical report, National Center for Geographic Information and Analysis, CA, (1991).
24. Binding, C.: Implementing Archaeological Time Periods Using CIDOC CRM and SKOS. In Proceedings of the 7th Extended Semantic Web Conference (ESWC), pp. 273--287, (2010).
25. M. Doerr, M., Kritsotaki, A., Stead, S.: Thesauri of Historical Periods - A Proposal for Standardization. In Proceedings of CIDOC Conference, (2005).
26. Allen, J. F.: Maintaining Knowledge about Temporal Intervals. In Communication ACM, vol.26(11), pp. 832--843, (1983).
27. Witte, R., Krestel, R., Kappler, T., Lockemann, P.C.: Converting a Historical Architecture Encyclopedia into a Semantic Knowledge Base. In IEEE Intelligent Systems, vol.25(1), pp.58--67, (2010).
28. Reutelshoefer, J., Lemmerich, F., Baumeister, J., Wintjes, J., Haas, L.: Taking OWL to Athens: Semantic Web technology takes Ancient Greek history to students. In Proceedings of the 7th Extended Semantic Web Conference (ESWC), pp. 333--347, (2010).
29. Bonomi, A., Mantegari, G., Vizzari, G.: A Framework for Ontological Description of Archaeological Scientific Publications. In SWAP, (2006).
30. Bonomi, A., Mosca, A., Palmonari, M., Vizzari, G.: Integrating a Wiki in an Ontology Driven Web Site: Approach, Architecture and Application in the Archaeological Domain. In Proceedings of the Third Workshop on Semantic Wikis: The Wiki Way of Semantics (SemWiki), (2008).
31. Mantegari, G., Palmonari, M., Vizzari, G.: Rapid Prototyping a Semantic Web Application for Cultural Heritage: The Case of MANTIC. In Proceedings of the 7th Extended Semantic Web Conference (ESWC), pp. 406--410, (2010).
32. Leclercq, E., Savonnet, M.: Access and annotation of archaelogical corpus via a semantic wiki. In Fifth Workshop on Semantic Wikis - Linking Data and People (Semwiki), (2010).