

Annotating Digital Libraries and Electronic Editions in a Collaborative and Semantic Perspective

Michele Barbera¹, Federico Meschini², Christian Morbidoni³, and Francesca Tomasi⁴

¹ Net7, Pisa, Italy, SpazioDati, Trento, Italy

barbera@netseven.it, barbera@spaziodati.eu

² Department of Humanities, Communication and Tourism, Tuscia University, Italy

fmeschini@unitus.it

³ Samedia Group, Università Politecnica delle Marche, Ancona, Italy

christian.morbidoni@gmail.com

³ Department of Classical Philology and Italian Studies, University of Bologna, Italy

francesca.tomasi@unibo.it

Abstract. The distinction between digital libraries and electronic editions is becoming more and more subtle. The practice of annotation represents a point of convergence of two only apparently separated worlds. The aim of this paper is to present a model of collaborative semantic annotation of texts (SemLib project), suggesting a system that find in Semantic Web and Linked Data the solution technologies for enabling structured semantic annotation, also in the field of electronic editions in Digital Humanities domain. The main purpose of SemLib is to develop an application so to make easy for developers the integration of annotation software in digital libraries, which are different both for technical implementations and managed contents, and provide to users, indifferently from their cultural backgrounds, a simple system which could be used as a front-end. We present, for this purpose, a final example of semantic annotation in a specific context: a digital edition of a literary text and the issues that an annotation task involves.

Keywords: ontologies, Open Collaboration, Linked Data, TEI, RDF.

1 Introduction

In the Library of Alexandria, the distinction between philologists and librarians was almost not existent, since the functions of acquisition, cataloguing and preservation were strictly related to an editorial work which main aim was to give to the texts the best possible rigour and accuracy [1]. The progressive specialization of skills has brought, as a natural consequence, the loss of a global view, also in strongly linked sectors like the ones quoted above. Let's only think about the different meanings that the term bibliography can assume when used with different prefixes such as 'analytic' or 'descriptive'. This same loss has been denounced by Vannevar Bush [2] and therefore his vision of the Memex was a possible solution for this situation. It is therefore quite a paradox that in digital information systems and frameworks published on the World Wide Web, representing on one side Digital Libraries (DLs) and, on the other,

Electronic or Digital Editions, this difference is, if possible, even sharper, having generated two different scholarly communities, which, even though overlapping, presents their own distinctive features.

In fact, by concentrating on cataloguing and digitizing collections rather than analysing the content of individual items in a collection, the DL is mostly focused on publishing mechanisms (or to better say the dissemination ones) — an opposite extreme from the focus of electronic editions on textual encoding, in particular the one based on the TEI standard [3]. The DL, instead, is by nature agnostic towards the contents it has to manage, since they could be very heterogeneous, preferring therefore to ignore the granularity of an encoded text, which is fundamental for an electronic edition. Moreover, a DL contains reproduction of physical objects or content *born-digital*, while in the current situation the primary sources upon which an electronic edition is based have an almost and exclusively analogical origin. But interaction between these two paradigms is actually taking place, even though, in a not very organized and co-ordinated way, following what clearly are physiological patterns, but in a deeper and more dynamic modality than the one allowed by the analogical dimension¹.

The electronic edition and the DL are modelled on the needs, experiences and uses of two different communities, even though related, and this is a natural expectation. What is not completely expected is the (re)definition of the respective natures, and overall modalities of interaction of these two entities, in particular from the point of view of their (re)modelling, following computational principles. In fact now the differences have a logical base rather than a physical one.

Even though the technologies are necessarily the same, this does not assure a complete homogeneity at the level of methodologies, approaches and solutions, in other words it does not guarantee the perfect overlapping and compatibility of the hypothetical semantic models. This discrepancy is mitigated by observing two opposed and complementary movements that are currently taking place. On one side the electronic editions and archives are expanding, becoming more and more complex and stratified, while on the other side DLs are becoming more and more granular (together with their natural tendency for progressive growth). In fact, of the two currently available formal models for DLs, the *5S* [7] and the *DELOS Reference Model* [8], the former, being directly based on first-order logic, is granular and expressive enough to model also electronic editions, both the textual encoding level and the very different relationships existing between the witnesses which made up the textual tradition.

Building on the facts presented until now, the most logical consequence is that a focus on semantic and formal models seems the only way for breaking the barriers between these DLs and electronic editions, even though, in this latter case the

¹ For instance, there is the progressive adoption of *IFLA FRBR* model, in projects such as *Perseus* <<http://www.perseus.tufts.edu>> [4] or the *Canonical Text Service* protocol <<http://chs75.chs.harvard.edu/projects/diginc/techpub/cts>> [5] or standards such as METS. On the digital libraries side, an interesting case is the use of the publishing framework *Cocoon*, inside *Dspace*, in a component called first *Manakin* and then *XMLUI*, used to customize the user interface. The *XML* schema used to transport the data, *DRI Schema*, is based on *TEI* for what concerns the actual contents [6].

prevailing diffusion of the TEI encoding, notwithstanding its many advantages, is a potential hindrance, since it brings the focus mostly on data structures which lack a formal semantics [9], [10].

The annotation task becomes then crucial. In the context of both DLs and electronic editing, the term annotation indicates the process of adding some kind of information to an existing digital resource. It can be a tag, a comment or some kind of structured metadata. The task of librarians is to provide sets of high quality annotations for each library resource, in order to help organizing the knowledge gathered in each library. These metadata are usually designed in accordance with standard library science practices and are meant to facilitate knowledge discovery by the generic library user. A DL can be accessed by several user communities, each one with a specific vision of the world and each one interested in a specific topic or aspect of knowledge. On the contrary, built-in metadata in DLs are often generic information (e.g. year of publication, author, historical period, artistic wave, etc.), or reflects a single viewpoint. They do not capture all the aspects the users might be interested in, thus being often of poor value with respect to interesting resources discovery.

A similar problem exists in more specialized types of digital publications, such as critical text editions. In digital editions it is common practice that the editors enrich the text by annotating it with TEI or other forms of markup. These annotations are then used to deliver a richer reading and searching experience for the final user. Despite this practice has certainly proved useful, it also suffers from a similar limitation to the one outlined above for generic DLs. Annotations made by the editors are intrinsically static and relative to a particular view of the world or school of thought. Differently than paper artifacts, digital resources can be easily exploited as social objects around which communities can collaboratively and continuously enrich the digital artifacts with different interpretations.

According to authoritative studies [11], [12], [13] DLs, and more specific collections of digital objects, should allow their users to annotate resources and leave comments. They should also let users share their index and classification schemata with other users [14]. We believe that DLs and more specialized collections in the Digital Humanities field, can greatly benefit from the availability of Web annotation tools based on Semantic Web and Linked Data technologies and the aim of our research is to show in which way.

The paper is structured as follows. Section 2 presents related works in the field of annotation. Section 3 discusses the approach taken within the context of the SemLib project, whose prototypal intermediate results (the project will publicly release its final results in December 2012) represent an interesting experiment in this direction.

Section 4 gives a specific case study of semantic annotation in a digital edition.

2 Related Works

Annotating Web documents like Web pages, part of Web pages, images, audios and videos is one of the most spread technique to create interconnected and structured metadata on the Web. In the last years several automatic, semi-automatic and manual

systems have been proposed that provide support for creating annotations at different levels and in diverse scenarios. Some applications have been developed as extensions of social bookmarking tools and have become a popular service over the Web with application as Delicious² or StumbleUpon³ that count millions of registered users. Other tools have been more specifically conceived for manually creating and sharing annotations in specific domains, including Digital Humanities and Cultural Heritage [15], [16]. Early implementations of manual annotation tools have been mostly developed as desktop applications or browser plugins (such as Zotero⁴ and others). With the growing availability of powerful client side Web programming tools and techniques, annotation tools then evolved in fully fledged Web applications such as EuropeanConnect Media Annotation Prototype [16] based on Annotea [17], One Click Annotator [18], the Open Knowledge Foundation's Annotator project⁵, SharedCopy⁶, A.nnotate⁷ and many others. Another widely applied method to create annotations is to use automatic and semi-automatic tools based on euristics like natural language processing, image recognition, audio and video segmentation. For textual content there are several widespread commercial services that automatically perform a light type of annotation known as entity extraction with a constantly improving degree of relevance (e.g. Opencalais⁸, Zemanta⁹, AlchemyAPI¹⁰).

While some of the existing tools address ease of use and wide adoption, they hardly provide support for expressing non trivial semantics, as establishing precise (typed) relations among digital objects or referring to specific entries in domain thesauri and vocabularies. In Semlib, the goal is to build different annotation GUIs to address different levels of expressivity, from simple tags to structured conceptual graphs, carefully balancing ease of use and expressivity. The other idea behind SemLib is that of representing annotations (simple or complex ones) in a uniform way (as RDF graphs), and expose them via REST APIs so to enable effective reuse of collaboratively created knowledge, for example to further enrich DLs.

3 Collaborative Semantic Annotation of Texts: The SemLib Approach

One of the main goals of the SemLib project [19] is the design and implementation of a semantic aware annotation system that can be easily used in conjunction with different DLs, requiring as less modification as possible to the existing DL software infrastructures, and that can be flexible enough to address different needs of specific

² <http://delicious.com/>

³ <http://www.stumbleupon.com/>

⁴ <http://www.zotero.org>

⁵ <http://okfn.org/projects/annotator/>

⁶ <http://www.sharedcopy.com/>

⁷ <http://www.a.nnotate.com/>

⁸ <http://www.opencalais.com/>

⁹ <http://www.zemanta.com/>

¹⁰ <http://www.alchemyapi.com/>

communities. Such differences both reside in required expressivity and complexity of annotations, which might range from simple tags to non trivial semantic relations among media content and other kind of entities, and in the use of different domain dependent terminology and vocabularies.

A core requirement in SemLib is that of enabling reuse of annotations, for example to leverage them as a crowd-sourced structured knowledge that might be used to enrich DLs themselves. While interoperability at data representation level is certainly a key feature with respect to this goal, the system has to provide effective ways to meaningfully consume such data, for example allowing external applications to search annotations and to obtain “slices” of the overall annotations (e.g. obtaining annotations from trusted users only, or those that involves relevant resources only, etc.).

3.1 From Tagging to Semantically Structured Annotations

The simple form of annotation, widely understood and adopted by the majority of Web users, is tagging. Keywords based tagging, however, has several disadvantages (between them: no explicit meaning and explanation; polysemy; synonymy; base form variation; specificity gap; reused in different systems). These poor semantics expressed by “traditional” tags prevents in fact the use of annotations to produce reusable structured knowledge, which is the core goal of SemLib. To overcome such limitations, SemLib supports “semantic tagging”, where each tag corresponds to an entry in a controlled vocabulary or ontology and it is a Web resource in itself, thus being resolvable into a natural language description by dereferencing its URL. A similar approach has been already experimented in the Common Tag initiative (Commontag.org). The current prototype allows users to transparently search for entities (semantic tags) in Freebase.com, providing auto-complete suggestions and resulting in external web resources to be associated to text fragments or pictures in a web page.

Interestingly, such web resource happens to be Linked Data sources. This means that they can be used to retrieve further information about the entities, allowing external applications, which consume such annotations, to immediately use such additional data in intelligent ways. In addition, existing APIs, such as DBPedia spotlight, are used to suggest simple forms of automatic tagging.

SemLib also supports more advanced types of annotations that exploit all the expressive power of the RDF data model, which goes beyond simple semantic tagging. As an example, suppose Alice is a scholar studying Italian literature. She finds a DL with some interesting novels. While reading one of them, she highlights a paragraph and creates an annotation specifying that such a paragraph cites Alessandro Manzoni and that attempts to give a definition of “Historical novel”. Semantic tagging as described above, does not allow to specify the relation (e.g. cites, defines) between the text and the related entities (e.g. “Alessandro Manzoni” and “Historical novel”), which is needed to answer queries like “what are the paragraphs that cites a given author?” or “What definitions of historical novel does the system know?”

The SemLib annotation tool supports the creation of such complex annotations by allowing user to collect different kind of “items” (they can be terms from a vocabulary, web pages, or fragments of them, e.g. sentences and pictures) and then connect them via semantically typed relations, thus in fact creating a semantic graph. The annotation tool can be configured to use custom vocabularies to accommodate the

needs of different DLs. Vocabularies can be published on the web in a simple JSON based format and then ingested by the tool by simply specifying their URL.

Cross-references annotations are an interesting special case that is often required by scholars. It consists in establishing a semantic relations between two media fragments, such as a sentence and other one in a different page (possibly in a different DL), or a sentence and a specific region of an image. From a conceptual and data representation point of view, they are equivalent to other semantic relations, however they raise new challenges at the user interaction level. In SemLib such kind of annotations are made possible by allowing user to bookmark media fragments, to surf to other web pages and then to reuse such bookmarked items in annotations.

3.2 Data Model and API

The representation of semantic annotation is composed by two distinct parts: the annotation context and its semantic content. The first represents information such as the author of the annotation, while the latter represents the actual meaning or knowledge that the user wanted to express in the annotation. The data model is illustrated in Fig. 1 as an RDF graph.

Different RDF based data models for representing web annotations have been proposed in literature. In SemLib we decided to base on the Open Annotation Collaboration (OAC) data model [20]. The OAC ontology is used to represent annotations contexts, which specifies, through the *oac:hasTarget* property, the web resources involved in an annotation. An additional concept used in the SemLib model is that of notebook, which is an aggregation of annotations. Each user can have multiple notebooks, e.g. to group annotations pertaining to different tasks or contexts. Notebooks have a central role in the overall functioning of the system, as they constitute the granularity where user privileges are attached and annotations are shared among users.

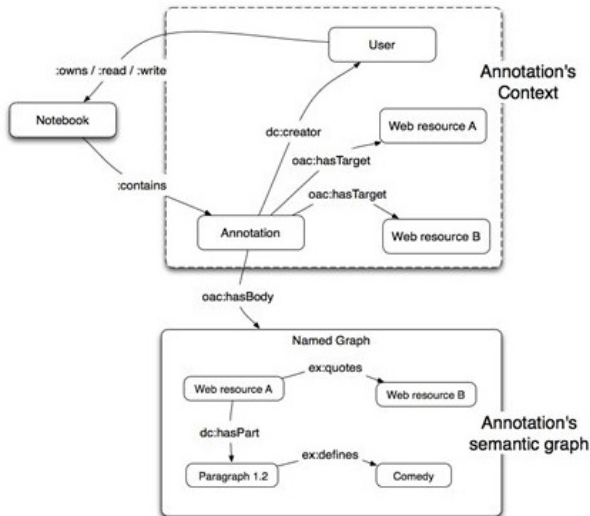


Fig. 1. The annotation data model

The *oac:hasBody* property links the annotation to its content. In SemLib the body, rather than being a text or a web page (as it happens in most of the examples given in the OAC specifications) is a RDF graph itself. The content graph is made addressable by using named graphs. As shown in Figure 1, a named graph is used as value of the *oac:hasBody* property. As annotations are stored in a quad-store and the SPARQL standard query language natively supports named graphs, such an approach results in more flexibility in consuming annotations.

Named graphs are sub-graphs of an RDF graph that can be merged and queried as single graphs. The use of named graphs allows, for example, querying only those graphs that belongs to annotations in a given notebook. The same can be done for collections of annotations grouped via other criteria (e.g. from the same user, involving the same resource, etc.). A set of RESTful APIs are exposed by the annotation server to provide an easy way of consuming slices of the overall annotations set in various RDF serialization formats, and additional custom queries can be performed via standard SPARQL endpoint.

While the annotation context is represented using a fixed ontology (an extension of the OAC ontology), semantic content of annotations can use any ontology that a certain domain requires (e.g. a TEI-derived ontology in the Digital Humanities main¹¹).

3.3 Enabling Collaboration in the Digital Humanities

One of the most relevant limitations of annotation systems based on embedded markup, such as HTML or XML, is the tight coupling between the annotation and the annotated object [21]. In the Digital Humanities community the problem is acute, partly because most of the times digital edition projects do not publish the XML annotated version of the text, but rather a derivative HTML or PDF version produced from the original XML. In these cases annotation ceases to be metadata to become part of the digital objects, thus not being reusable, preventing collaboration and further enrichments of the text.

Even when the XML source files are distributed along with the human readable version, they are seldom reused or integrated outside the boundaries of the systems in which they were originated. The semantics of the annotations is in most cases based on local interpretations and local extensions of the core vocabulary. Additionally, vocabularies and thesauri are not shared and the semantics of local extensions is not machine-readable. As a result, textual resources, their interpretations and enrichments, remain siloed within the boundaries of individual projects. Data is not shared, derived information and interpretation is hidden within the browsing applications and not addressable or reusable. As a result, digital collaboration practices are rather weak if compared to other disciplines where information sharing and reuse is more common (e.g. ‘hard’ sciences).

Stand-off semantic annotations based on the RDF standard, as described in the previous section, have the potential to overcome many of these limitations.

¹¹ TEI ontologies SIG <http://www.tei-c.org/SIG/Ontologies>

Annotations are useful in organizing and adding information to digital content, supporting single users in studying and exploring online resources. However, a lot more value can be added if annotations are shared with others, enabling, for example, virtual communities to perform collaborative tasks. In SemLib, both annotations and vocabularies can be published on the Web in machine-readable format (RDF) so that any authorized application can interpret annotations. This basically decouples the applications used to access the DL from the content itself, allowing for multiple “views or interpretations” of the same content to be published in a decentralized way. Another interesting side effect of the Semantic Web approach, is that it naturally enables collaboration: as long as users have the possibility to upload their annotation somewhere on the Web, they are free to keep enriching the content stored on the DL without the need of any coordination with the content holder. Users communities can share their work by simply exchanging URLs that point to their annotation graphs. At a later stage, user annotations can also be made authoritative by the library curators, by incorporating them into the DL. Annotations are provided with provenance meta-data, so that it is always possible to determine who made a specific annotation and when.

While the prototype has not yet been released to the public, some online screen-casts demonstrate the core functionalities and user interactions¹².

4 Adding Meaning. A Case Study

The case study we present here is a digital edition of a XV century collection of letters now held in different libraries and, mostly, archives. The purpose of the edition is to experiment a concrete case of semantic annotation starting from a sequence of XML/TEI files, regarding the same field.

The correspondence documents the professional relationships managed by the Florentine librarian and copyist Vespasiano da Bisticci, who was also the leader of a school of copyists, maker of some European libraries’ manuscript repositories. The correspondence is with notable people of that period and the content regards mostly the trade of manuscripts copied, proposed or requested by/to Vespasiano. A lot of these manuscripts had been identified in codices now held in various libraries in all Europe. From the letters we can learn about features of these manuscripts: the materials, the copyists, the costs, but also the names of latin, greek and humanistic authors and texts that were the most fashionable at the time.

The purpose of the digital edition, moving toward a DL of digital objects, is on one side to represent the information that is implicitly connected inside the source (people with manuscripts, manuscript with lexicon); on the other to create semantic links between the information inside the letters and correlated Web resources, useful in order to describe this information (people, manuscripts, lexicon). The first purpose is thus to create relationships between people and manuscripts related to people at some level

¹² http://www.youtube.com/watch?v=gVA_v152Qn0,
<http://www.youtube.com/watch?v=z1hXr5K3kTM>

(the copyist, the owner, the requester) and the manuscripts with the lexicon used in order to describe them; this information born naturally from the letters and have to be expressed with semantic assertions through external annotations. The second aim is to create relationships between the specific document and other documents: people and resources useful to describe people (public prosopography), mentioned manuscripts with the existent codices (catalogues of manuscripts), the lexicon with repositories of technical words (thesauri) using the Linked Data system.

4.1 The Embedded Markup

The markup model we present covers all the different aspects of the edition, that is the model reflects the different access points to the letters' content. The focus is on: persons mentioned in the letters; manuscripts realized by Vespasiano's school; the technical lexicon of the copy and of librarian trade.

At the markup level it is quite easy to represent, with the appropriate TEI elements, all this information. The base TEI markup let us identify: proper name (`<persname>` and `<placename>`) and referring string (`<rs>` with `@type` for specification), mentioned manuscripts (`<bibl>`) and related author (`<author>`) and title (`<title>`), technical term (`<term>`) on various field (`@type`).

The `@ref` value in the instances of `<persname>`, `<placename>` and `<rs>` allows a first identification of individuals being mentioned: missing parts of names are solved outside and variants are associated to the same instance. The same attribute `@ref` was also used for the element `<bibl>` and `<term>`, with a TEI customization. The `@ref` value points to a specification of the item stores in a place outside the document. Therefore, within these elements, an access point is defined as an element owning a URI reference that points to `@rdf:about` attribute in the external repository. In this way each pertinent string of characters or fragment has an URI.

If the URI pointing gives us the possibility to formally describe annotated elements in the external representation, we need a system to create connections between these annotated elements. These connections must answer to questions like: which relation exists between a person and a manuscript? And which one between the same manuscript and a technical word used to describe the manuscript? But we need to answer also to questions like: which exemplar of the manuscript has been realized? Is it still available in any library in Europe? What a specific technical word means? There are any other occurrences in other repositories? Who is a mentioned person?

For the first set of questions we need some more formal semantics in order to describe relationships between annotated data elements. For the second set of questions we have to use techniques useful in order to connect the edition with existent resources on the Web.

4.2 The External Information

Once the texts are annotated, and each pertinent string of characters has an URI, it is necessary firstly to focus on what kind of additional information it is possible to define in the separated documents. Making each occurrence (persons, manuscripts and

lexicon) accessible via URI-based pointing we can create a public authority list. Starting from Vespasiano's letters it became possible to expose authority information about people, manuscripts, technical lexicon regarding XV century culture, in order to start to create an open and public authority list, to be integrated in authority records, as a set of Web resources.

We need firstly to decide which kind of relevant added data it is possible to specify for each of the three categories. Then we have to reflect on which kind of relationships we could define, expressing the content thought stand-off semantic annotations.

Persons. At this level we define relationships: with other persons; with places; with dates; with other resources (like multimedia data); and with mentioned manuscripts (i.e. a person could be owner of a codex created by Vespasiano's school).

Manuscripts. With regard to manuscripts mentioned in the letters it is possible to create link to other repositories and establish relationships with both persons and terms. We have to create relationships: with the codex repository; with the codicological description; with a digital image; with the digitalized full text; with a person relevant to it (i.e., the owner, the requester, the copyist, etc.).

Lexicon. The analysis of the technical lexicon used by Vespasiano is an interesting exploration in the history of the book and in the actual trade of the copy, and it is sometimes connected to the manuscripts realized. At this level it is possible to define relationships between manuscripts and lexicon used.

4.3 The Knowledge Base

All this external information could be described in a formal way through RDF assertions, like similar researches did (i.e. [22], [23]) and the annotations could be accessible over the Web as Linked Data sets. Mostly we need to focus on the fact that between these concepts (persons and manuscripts; manuscripts and lexicon) we can define relationships that provide greater conceptual depth and that can be easily expressed in a formal language, creating a good model for the representation of the content of the letters: we can establish unambiguously the different relationships existing between a person, a manuscript and a term, such as a person being the owner, the copyist, or the client of a codex and this manuscript is described with a specific term (p.e. Piero de' Medici is owner of Plutarco's Vite, which are realized in "chordovani"). Mapping the specific created class (persons, manuscripts, lexicon and the different kind of relationships) with predicates defined in existent model a suitable ontology could be distributed. We are now analyzing different predicates for internal connection, starting from these consideration: a person could be *owner-of*, *copyist-of*, *illuminator-of*, *requester-of* a manuscript; a manuscript could be *created-for*, *requested-by*, *copied-by*, *illuminated-by* a person / *described-with* terms; the lexicon is *related-to* a manuscript.

But like we explain in the previous section the external information could be linked to some external resources. We are studying some predicates. For person, TEI prosopography integrated with CIDOC-CRM¹³ for relationships between Agents, Physical things, Events and Places. For manuscripts an ontology properly extending and

¹³ CIDOC Conceptual Reference Model, <http://www.cidoc-crm.org/>

customizing the FRBR¹⁴ can be used to capture the subtle difference between the physical codex, the content it has and the work the codex is a *manifestation-of*. For the lexicon SKOS¹⁵ provides a basic ontological foundation for a terminological thesaurus and the relationships of the terms defined therein and elsewhere.

But we mostly need to reflect on the fact that the mentioned manuscripts are specific codices that can be found nowadays in a specific library, or a person has an iconographic representation that could be found in a certain cultural institute. With URI references and RDF representation we started to create relationships between people, manuscripts and lexicon of the letters and the related concept all over the Web, using existent Linked Data sets and exposing our annotation as Linked Data sets in open collaboration.

5 Conclusions

Having started from a general description of current epistemological and ontological differences between digital libraries and electronic editions, together with the related underlying rationale, the choice of focusing on open and semantic annotation has been considered by the authors a strategic one, since as it has been demonstrated this “functional primitive” has the potential to be a bridge between these two different but strongly related worlds. Therefore both the act of annotating a text, thus giving it the status of “scholarly” and the addition of (meta)information to a generic resource can greatly benefit from the adoption of a common formal model, which at the same time make explicit both their actual natures and potentialities.

The aim of SemLib project is to use standard technologies to create a really usable application, easy to use for every kind of user and easy to integrate in heterogeneous digital libraries, therefore filling up a current gap in the landscape of annotation applications on three different levels: actually usability, ease of use, integration. For these reasons we are organizing focus groups which would consolidate the theoretic foundations which led the implementation of the application.

Acknowledgements. The research leading to these results has received funding from the European Union's Seventh Framework Programme managed by REA-Research Executive Agency ([FP7/2007-2013][FP7/2007-2011]) under grant agreement n. 262301.

References

1. Ore, E.S.: ... they hid their books underground. In: Deegan, M., Sutherland, K. (eds.) Text Editing, Print and the Digital World, pp. 113–125. Ashgate, Farnham (2009)
2. Bush, V.: As We May Think. Atlantic Monthly (1945)
3. Burnard, L., Bauman, S. (eds.): TEI P5: Guidelines for Electronic Text Encoding and Interchange. TEI Consortium (2007)

¹⁴ Functional Requirements for Bibliographic Records, <http://www.frbr.org/>

¹⁵ Simple Knowledge Organization System,
<http://www.w3.org/TR/skos-reference>

4. Mimno, D., Crane, G., Jones, A.: Hierarchical Catalog Records: Implementing a FRBR Catalog. *D-Lib Magazine* 11(10) (2005)
5. Smith, N.: Citation in Classical Studies. *Digital Humanities Quarterly* 3(1) (2009)
6. Phillips, S., Green, C., et al.: Manakin: A New Face for DSpace. *D-Lib Magazine* 13 (2007)
7. Gonçalves, M.A.: Streams, Structures, Spaces, Scenarios, and Societies (5S): A Formal Digital Library Framework and Its Applications. *Computer Science and Applications*. Virginia Polytechnic Institute and State University, Blacksburg, Virginia, U.S.A. PhD: VII-153 (2004)
8. Candela, L., Castelli, D., et al.: The DELOS Digital Library Reference Model. *Foundations for Digital Libraries Version 0.98, DELOS Network of Excellence on Digital Libraries* (2007)
9. Raymond, D.R., Tompa, F.W., Wood, D.: Markup Reconsidered. In: *First International Workshop on Principles of Document Processing*, Washington, DC (1992)
10. Buzzetti, D.: Rappresentazione digitale e modello del testo. In: *Il Ruolo Del Modello Nella Scienza e Nel Sapere*, pp. 127–161. *Accademia Nazionale dei Lincei*, Roma (1998)
11. Agosti, M., Ferro, N.: Annotations: Enriching a Digital Library. In: Koch, T., Sølvberg, I.T. (eds.) *ECDL 2003. LNCS*, vol. 2769, pp. 88–100. Springer, Heidelberg (2003)
12. Agosti, M., Ferro, N., Frommholz, I., Thiel, U.: Annotations in Digital Libraries and Collaboratories – Facets, Models and Usage. In: Heery, R., Lyon, L. (eds.) *ECDL 2004. LNCS*, vol. 3232, pp. 244–255. Springer, Heidelberg (2004)
13. Agosti, M., Ferro, N.: An Information Service Architecture for Annotations. In: *DELOS Workshop Digital Library Architectures*, pp. 115–126 (2004)
14. Kruk, R., Decker, S.: Semantic Social Collaborative Filtering with FOAFRealm. In: Decker, S., Park, J., Quan, D., Sauermann, L. (eds.) *Proc. of Semantic Desktop Workshop at the ISWC*, Galway, Ireland, November 6, vol. 175 (2005)
15. Bradley, J.: Pliny: A model for digital support of scholarship. *Journal of Digital Information (JoDI)* 9(1) (2008)
16. Haslhofer, B., Momeni, M., et al.: Augmenting Europeana Content with Linked Data Resources. In: *6th International Conference on Semantic Systems* (2010)
17. Kahan, J., Koivunen, M.R.: Annotea: An Open RDF Infrastructure for Shared Web Annotations. In: *Proceedings of the 10th International Conference on World Wide Web*, pp. 623–632 (2001)
18. Heese, R., Luczak-Rösch, M., et al.: One Click Annotation. In: Williams, G.T., Grimnes, G.A. (eds.) *CEUR Workshop Proceedings*, vol. 699 (February 2010)
19. Morbidoni, C., Grassi, M., et al.: Introducing SemLib Project: Semantic Web Tools for Digital Libraries. In: *International Workshop on Semantic Digital Archives, TPD, Berlin, Germany, September 29* (2011)
20. Sanderson, R., Van de Sompel, H. (eds.): *Open Annotation: Alpha3 Data Model Guide* (October 15, 2010), <http://www.openannotation.org/spec/alpha3/>
21. Thompson, H.S., McKelvie, D.: Hyperlink semantics for standoff markup of read-only documents. In: *SGML Europe* (1997)
22. Vieira, J.M., Ciula, A.: Implementing an RDF/OWL Ontology on Henry the III Fine Rolls. In: *OWLED 2007*, Innsbruck, Austria (June 2007)
23. Tummarello, G., Morbidoni, C., Pierazzo, E.: Toward Textual Encoding Based on RDF. In: *9th ICCCE International Conference on Electronic Publishing*, Leuven-Heverlee, Belgium (June 2005)