A Semantic Linking System for Canonical References to Electronic Corpora

Matteo Romanello

Ca' Foscari University of Venice

1. Introduction

Recently on the Web the number of on-line free accessible corpora of ancient languages has appreciably increased. Most of the time existing projects aimed at building electronic corpora of ancient languages make use of an internal linking system which, on one hand allows for a worthy degree of hypertextuality to be reached but also it can produce a fairly closed system of hard-linked resources.
And so a shared standard or a well established "best practice" to determine how to encode references to texts stored in any given corpus, within an (X)HTML document does not yet exist. Microformats, one of the most cutting-edge technologies of the Web 2.0, could be applied to adequatly fill this gap.
Therefore this paper begins by describing the architecture of a digital library of classical literature texts where the coupling between distributed corpora of primary sources and a great variety of secondary sources is realized by adding a layer of semantic data to textual references.
Finally, the paper ends by proposing a possible implementation of a linking system where the aggregation of relevant information is achieved by implementing client-side software. This software should be capable of understanding appropriately encoded references within texts and to resolve them in a context sensible way. The data and information of different kinds obtained this way are supplied from a open-ended number of content providers available as Web Services.

2. Classical Digital Library Architecture

For the information architects involved, switching from a printed to a digital library means finding new constructive paradigms to try to take advantage of the Net's properties such as connectivity and decentralization.
Indeed in a network environment it becomes possible either to build a universal distributed library where different libraries are avoided from having those duplicated texts that are inevitable in a printed context, either providing to readers a context-sensible reference linking system, allowing to move directly from textual references to relevant resources. The following example shows how it works: if reference linking is possible and a distributed library of primary sources is available, a reader could

move from a citation like Hom. *Il.* 1.1 to all the available editions or translations, comparing critical editions or finding related resources (illustrations on pottery, reviews of articles or books concerned with that specific text passage etc.). Furthermore, a digital library with a distributed architecture seems to be an infinitely more suitable model than a galaxy of isolated and un-interoperable digital libraries growing with each single projects or research grants.
The following section will highlight the main issues concerning the overall architecture of post-incunabular digital libraries [Crane et al. 2006] and their technological solutions. Besides special attention will be paid to digital libraries within the classical literature studies field.

2.1 Electronic secondary sources: state of the art

In ancient language studies, the importance of primary and secondary sources is so significant that scholars' research work can be seen as lying for the most part in creating, retrieving or modifying links between these two different kinds of sources. Indeed when they compare texts in order to produce new interpretations, they constantly draw new ties among primary sources, and they create new ties thus providing secondary sources in support of their thesis.
While primary sources are mostly organized as electronic corpora, texts under the definition of secondary sources comprise a more heterogeneous variety of resources, such as monographies and research articles, reviews of resources or specialist blogs all published electronically. Currently, secondary sources (particularly academic publications) do not overcome an incunabular stage [Crane et al. 2006]. A few on-line publications are provided as HTML, whereas most of the journals just keep publishing a binary file reproducing exactly the printed issue.
In these publications, semantic data about authors and text references are mostly not encoded, though in a few cases these are marked up as simple links to other stable on-line resources (e.g. Perseus' on-line texts),[1] following an internal linking system that varies from project to project. To date, digital canonical text references still simply replicate printed ones, where abbreviations and implicit statements constantly require human disambiguation capabilities. Besides they are treated by robots and information retrieval tools merely as strings without the slightest semantic markup that instead would permit an automatic and metadata-aware information processing.

2.2 Evolution of electronic corpora

The characteristics of the projects started in the past concerning building electronic corpora are so strictly dependent on the available technologies, that they can be thought in terms of diachronic evolution.

1 The Perseus Digital Library, `http://www.perseus.tufts.edu`

Indeed, as pointed out by Neel Smith [2004], "the TLG founded in 1970s is a response to the potential of mass storage and rapid retrieval in digital media" and "the Perseus project founded in 1980s is a response to richer media and higher-level data structures".
Currently, a distributed digital library built upon various web-oriented protocols and standards would probably be the response to the technological convergence of XML-related technologies, an easier web service communication over interfaces designed in conformity with a Representational State Transfer (REST) architectural style [Fielding and Taylor 2002] and a distributed architecture encouraged by the web itself.

2.3 Building distributed corpora: the CTS protocol

Now there is a recent protocol which allows different digital repositories of TEI-encoded texts to join together, providing a common protocol to make these collections interacting with each other: the Canonical Text Service protocol (CTS),[2] developed among the others by Neel Smith and Chris Blackwell at Harvard's Centre for Hellenic Studies.
The protocol relies on the conceptual model described by the Functional Requirementes for Bibliographic Records (FRBR), distinguishing between a work and its different exemplars, while also introducing some slight differences, such as introducing the concept of 'workgroup' to replace the 'author' one. One of its great features is that the protocol allows to be reached a higher granularity accessing documents hierarchically and supports the use of a citation scheme referring to each level of the entire document hierarchical structure.
CTS uses Uniform Resource Names (URN), called CTS-URNs, to identify univocally authors and works within corpora of canonical texts. Although to date CTS has been applied only to classical texts, it could be suitable for all discrete collections of TEI-encoded texts. Furthermore it is built upon another important protocol, the Registry Services Protocol,[3] providing "an automated interface to authority lists". The distributed and wide scope nature of Registry Services protocol allows to be created several lists of identifiers publishing them as Registries, that may be used, for instance, as unique identifiers for authors and works within collections of texts accessible via the CTS protocol.
Finally, someone could ask "why did you decide to build a linking system for primary and secondary sources upon the CTS protocol, although is not yet widely adopted?". The first reason is that it is effective and fits perfectly to scholars' needs when referring and accessing hierarchical sections of canonical works within a digital library. Secondly, it is built entirely upon common standards and responds to the Web's decentralised nature. Further, recently the Perseus Project started implementing

2 The Canonical Text Services (CTS) Protocol, current version: 1.1, http://katoptron.holycross.edu/cocoon/diginc/specs/cts

3 The Registry Services Protocol, current version: 1.0.rc1 , http://katoptron.holycross.edu/cocoon/diginc/specs/registry

a CTS-compliant web interface for its Digital Library [Weaver 2007]. This choice of allowing potentially thousands of web applications to use its raw XML data, made by the biggest classical digital library project is destined to increase Perseus' popularity rather than reduce it and seems to bode well for a wide-scale adoption of the protocol itself .

3. The Big issue: Linking primary to secondary sources

The way scholars write and use such references in the printed context lacks any separation between content and presentational matters, because the printed medium itself suffers in the same way. Indeed it is not possible to embed within a given text reference (e. g. Hom. *Il.* 1.1), contained in a printed page, some information about its own implicit meaning.
Instead, a digital equivalent of printed references would allow these two different layers to be kept separate, achieving the goal of having for instance, the same semantic information formatted in different citation styles. However the big issue is that there is not yet a widely adopted, digital and equally precise equivalent to printed canonical text references. The aim of this paper is to solve this big issue in order to give scholars who use the Digital Library for their research activities the tools they deserve, from reference linking to targeted and semantic search.

3.1 Scenario

The desirable linking system between primary and secondary sources, whose implementation is the main goal of this paper, is well described by a scenario where the aggregation of information relevant to a given text reference is accomplished by a client-side application that is capable of understanding appropriately encoded references and where the function of retrieving information and supplying the user with content is left to external service providers who present at least one web service interface.
Further, this linking system is well described by a scenario where: 1) any encoded reference to corpora texts becomes a direct link to a number of other digital resources spread over the Web; 2) web agents discover new resources that concern a given canonical author or work (e.g. the author referred to in a text reference); 3) programmers of third-party applications enhance interoperability among web applications that use different data architectures and communication protocols; 4) users experience a rich navigation through different digital resources, starting with simple textual references; 5) structured semantic information is reusable.

3.2 Architectural approach

Currently (for the reason that on-line primary and secondary resources can be two tightly coupled systems or totally discoupled ones) it seems to be necessary to keep them in a loosely coupled system [Fielding & Taylor 2002].

Indeed, within existing Digital Library projects connections between primary and secondary sources (when they exist) are established by using an internal linking system which permits a worthy degree of hypertextuality to be reached but produces fairly closed systems of hardlinked resources.
Instead, if primary and secondary sources were joined together forming a loosely coupled system, it would be possible to resolve each link to an open-ended number of on-line resources. Furthermore, linking primary to secondary sources in a loosely coupled way implies that some semantic information about the texts referred to have to be embedded within documents published on the web in order to allow automated agents processing those references.

3.3 Properties of canonical text references

First of all it seems necessary to define the properties of references to canonical texts in order to find a technical solution that really fits scholars' needs and usual practices. Since any reference may or may not be written following a given citation scheme and implying some implicit information, it becomes difficult to operate a fully automated analysis of raw text in order to identify references to corpora texts and to extract semantic information from them.
Examining some use cases (see Fig. 1) it is possible to distinguish at least three types of references (abridged, unabridged, implicit) on the basis of their appearance and three types (to a canonical author, to a canonical work and to a precise text passage) on the basis of its meaning. In addition, since references are language-dependent also the language variable should be considered, that exponentially increases the range of possible equivalent references: Homerus is also referred to as Homer (eng.), Homère (fr.), Omero (it.) etc.
A common practice among scholars in classics is to refer to different editions by shortening the editor's name or referring to the current edition of a text just omitting the editor's name (e.g. Hom. *Il.* 1.1 is referred to the Allen's current edition). Indeed, a common practice is to leave the editor's name when a work is known entirely and to indicate it instead when it is known by fragments. In particular this practice makes references even more context-dependent since, when another edition becomes the current reference edition for scholars, the meaning of the reference itself changes. Consecutive references to different loci of the same texts are usually written specifying once the text referred to and then indicating just the line or the pointed chapter number by the author to.
In conclusion, from the analysed use cases the result is that the inclusion respectively, of the author and of the work referred to, should be considered at the same time as the zero degree and the minimum properties of a canonical text reference. The indication of text range and editor's name appear to be additional information that contributes to setting up a precise text reference, concerned with a specific exemplar (both digital or printed) of a canonical work.

1. *Politics*
2. like Aristotle claims
3. *Politics* of Aristotle
4. Artist. *Pol.* 1304B
5. Line 1 of the first book of Homer's *Iliad*
6. Hom. *Il.* I 1
7. A 1 (Upper-case Alpha 1 in the Hellenistic books notation is equivalent to Hom. *Il.* 1.1)

Figure 1: Some use cases for (more or less structured) canonical references.

3.4 Technical solutions

To date, many technologies that try to solve the problem of adding semantic information to web pages (either embedding or attaching it) are available, primarily the Resource Description Framework (RDF), a triple-based language built upon XML and concerned directly with ontologies and with the Semantic Web.

Despite their recent popularity, Microformats are not the only solution to embed metadata within (X)HTML tags. Other possible solutions are at least RDFa,[4] the Microformats standard competitor fully supported by W3C, and eRDF.[5] Since each of these technologies has its own benefits and drawbacks, choosing one of them would be in any case a trade-off between the optimum solution and the most easily feasible one.

Microformats (described in section 4 of this paper) are suitable for micro-metadata and are considered a sort of bottom-up path to the Semantic Web. Terms to describe this format appropriately are "lower-case semantic web" and "real world semantics", as coined by Khare and Çelik 2006. Indeed they are easier for humans to understand and to implement than Resource Description Framework or Web Ontology Language (OWL), while they are also significantly less powerful. However, the capability to extract virtual RDF graphs from non-RDF data through the GRDDL (Gleaning Resource Descriptions from Dialects of Languages) mechanism guarantees the forward-compatibility of Microformats, as well as of RDFa (Resource Description Framework attributes) and eRDF (embedded Resource Description Framework), using Semantic Web technologies such as RDF and OWL [Hausenblas et al. 2007].

Instead RDFa permits more complex encoding of metadata using some attributes from XHTML's meta and link elements and generalizing them to all elements. Data encoded with RDFa can be more easily RDF-ised using standard GRDDL transformations rather than those microformatted. Nevertheless there are some compatibility issues among RDFa and XHTML 1.1 that are destined to remain unresolved until the release of XHTML 2 [Graf 2007].

4 http://www.w3.org/2006/07/SWD/RDFa/syntax

5 http://getsemantic.com/wiki/ERDF

4. A Microformat vocabulary set for canonical text references

4.1 What are Microformats?

> *Designed for humans first and machines second, Microformats are a set of simple, open data formats built upon existing and widely adopted standards. Instead of throwing away what works today, Microformats intend to solve simpler problems first by adapting to current behaviors and usage patterns (e.g. XHTML, blogging).*[6]

Microformats are currently being developed by the developers community through a specific wiki, where it is possible to find principles and guidelines to be followed during the Microformats design process.
Types of data which are currently being encoded using Microformats include geographical data, personal profiles, relationships, bibliographical data, reviews and curricula vitae. The use of Microformats particularly in blogs and social networks, demonstrated how semantic data could be embedded successfully inside compounds of Plain Old Semantic Html (POSH) and without mixing content with presentational features, that are instead managed through Cascading Style Sheets (CSS).

4.2 Microformats suitability

With respect to the Microformats competitors' purposes, the complexity of data we deal with is not enough to require the capability of RDFa or eRDF to embed within XHTML elements and attributes such complex semantic data that have to be expressed by using multiple (several) RDF vocabularies.
Due to their rapid and widespread success, Microformats have undoubtably contributed to making more popular the topic of how to realize the Semantic Web and also to draw attention to it. Besides they appear an easy way to implement a solution that may be considered as a starting point toward the future adoption of more powerful and semantic web-oriented technologies, such as RDF.
Although it contrasts in some way with the Web's decentralised nature itself, the centralised development of new Microformats prevents redundancy and encourages discussion on standard proposals. Furthermore, it has been announced that both coming versions of two of the most popular web browsers (Internet Explorer 8 and Mozilla Firefox 3) will feature support for this technology.
Finally, the success of Microformats-based technologies is not restricted just to blogs and social networks. Librarians too are taking advantage of this approach by introducing OpenURL CoinS (ContextObject in SPAN),[7] a technique to embed citation metadata in html `<span>` elements in order to obtain a reference linking

6 http://microformats.org/about
7 http://www.ocoins.info

system by allowing processing agents to discover, process and reuse them. Indeed by processing such encoded citation metadata, agents can substitute them with context-sensitive links (i.e. OpenURLs), pointing to a referrer service capable of mapping those metadata with entries from libraries' OPACs.

4.3 Design of the proposed Microformat vocabulary set

After the properties of canonical text references have been singled out, it is necessary to design a Microformat vocabulary set in accordance with the above specified requirements, choosing the most suitable among POSH tags. The design process should conform to Microformats design principles, such as embeddability and modularity, in the way desired by developer community [Allsopp 2007].

The Microformat vocabulary structure and the terms used as values for some "class" attributes (e.g. "textgroup" or "projid") were derived from using the CTS protocol, which was assumed to be as a declared model to describe the hierarchical structure of canonical texts collections.

The first principle states that a "Microformat must solve a specific problem": our initial problem is how to link primary and secondary sources within a Digital distributed Library in a way open-ended, semantic and cross-language way. According to the fourth principle claiming that designing mciroformats means "paving the cowpath", the proposed Microformat does not change the current behaviour of the users for whom it is especially designed (i.e. scholars). Hence, our solution should fit with the actual practice of citing canonical texts outlined in section 3.3.

Here I propose three Microformats that may be combined to encode almost the entire range of possible references individuated above (the detailed structure of each Microformat is illustrated in Tables 1-3). In particular, the decision to split the vocabulary set into three different Microformats responds to needs of embeddability and modularity, while also corresponding to the classification of references by meaning adopted early. Therefore each Microformat is suitable to encode abridged, unabridged and implicit references as well.

The elements used to compose the Microformat vocabulary set are the most semantic among POSH elements. Since the entire canonical reference could be defined in a broad sense as a citation, a `<cite>` element was used as a container element. Similarly, to properly encode both the author's name and the title of the work referred to, when they are expressed in an abridged form, the use of an `<abbr>` element is preferable to that of a `<span>` element.

As highlighted by the chunk of code reproduced in Fig. 1, the value of the class attribute of each element indicates the property expressed by the element itself as implied by the fundamental Microformats technique aimed at adding pieces of semantic information to POSH elements.

```
<cite class="ctref">
  <span class="ctauthor">
        <span class="projid">urn:cts:greekLit:tlg0012</span>
        <abbr class="name" title="Homer">Hom. </abbr>
   </span>
   <span class="ctwork">
        <span class="projid">urn:cts:greekLit:tlg0012:tlg001<span>
        <abbr class="title" title="Iliad">Il. </abbr>
   </span>
   <abbr class="range" title="20.131-20.137">20.131-7</abbr>
   <span class="edition">
        <abbr class="description" title="Allen" />
   </span>
 </cite>
```

Figure 2: Example of a microformatted textual reference (Hom. *Il.* 20.131-137) pointing to the current edition of Homer's Iliad by T. W. Allen, where the edition statement is implicit for a work that is not known by fragments.

Then, in order to univocally identify authors and works referred to, a universal scheme of identifiers capable of describing a collection of texts by assigning some unique identifiers to them and to their bibliographic properties (author's name, title of work, edition etc.) becomes necessary. An opaque identifier is par excellence language-neutral and thus satisfies the above outlined requirements of a linking system joining together primary and secondary sources.
A Scheme of URNs suitable for our purposes and already existing is the authority list called CHS Canon of Greek Literary Works, produced within the above mentioned CTS project and accessible through the Registry Service protocol. The CTS URNs of both canonical author and work referred to in a given text reference are therefore embedded in the proposed Microformat vocabulary set within a dedicated `<span>` element.
Finally, the forward-compatibility with RDF can be achieved through the use of some already existent vocabularies and ontologies, such as Expression of Core FRBR Concepts in RDF,[8] Citation Oriented Bibliographic Vocabulary[9] and Dublin Core Metadata Element Set.[10] Thus microformatted references can be extracted as RDF triples by using these vocabularies to express semantic data, after have been processed by a GRDDL transformation.

8 http://vocab.org/frbr/core
9 http://vocab.org/biblio
10 http://purl.org/dc/elements/1.1

POSH Element	*Class Attribute Value*	*Class and properties (OOP notation)*	*Content*	*Location*	*Data Type*
`cite`	`ctref`	Reference	—	Root element	Microformat
`abbr`	`range`	Reference.range	Expanded range	Element content	string
			Abridged range	*Title* attribute	string
`span`	`edition`	Reference.edition	—	Child elements	html
`span`	`projid`	Reference.edition.-projid	Edition identifier	Element content	CTS URN
`abbr`	`description`	Reference.edition.-description	Expanded edition statement	`<abbr>` *title* attribute	string
			Abridged edition statement	Element content	string

Table 1: Proposed Microformat for references to a canonical text passage (ctref).

POSH Element	Class *Attribute Value*	*Class and properties (OOP notation)*	*Content*	*Location*	*Data Type*
`span`	`ctauthor`	Author	—	Root element	Microformat
`span`	`projid`	Author.projid	Author identifier	Element content	CTS URN
`abbr|span`	`name`	Author.name	Abridged author's name	Element content	string
			Expanded author's name	*Title* attribute	string

Table 2: Proposed Microformat for references to a canonical author.

POSH Element	Class *Attribute Value*	*Class and properties (OOP notation)*	*Content*	*Location*	*Data Type*
`span`	`ctauthor`	Work	—	Root element	Microformat
`span`	`projid`	Work.projid	Work identifier	Element content	CTS URN
`abbr|span`	`title`	Work.title	Abridged work title	Element content	string
			Expanded work title	*Title* attribute	string

Table 3: Proposed Microformat for references to a canonical work.

4.4 Benefits of microformatted canonical text references

To sum up, a key concept of the proposed linking system is the separating cut of the layer of references within web documents containing embedded metadata, the layer of applications or agents capable of understanding such references and the layer of provided services. These distinctions should ensure both the open-ended nature and the interoperability of the entire system.

Therefore, possible services for taking advantage of such microformatted references include services of reference linking and targeted search, information aggregators, mashups and applications which allow the reuse of data extracted from those references.

An engine for targeted search, modelled on existing services that allow for searching among information tagged using a given Microformat (e.g. Technorati's Kitchen[11] supporting in particular the search among hCard,[12] hCalendar[13] and hReview[14] Microformats), will make possible the retrieval of a huge number of resources relating to a given author or work, giving the scholars' information retrieval tools infinitely more precision than the usual web search engines.

Furthermore, applications allowing reuse within desktop applications of microformatted data extracted from web pages may be built. It is possible to figure out an application which should allow scholars to manage collections of the most frequent canonical text references and to export them formatted according to a given citation style (or in a localized or abridged, rather than unabridged, form) to a word processor.

11 `http://kitchen.technorati.com`
12 `http://microformats.org/wiki/hcard`
13 `http://microformats.org/wiki/hcalendar`
14 `http://microformats.org/wiki/hreview`

But for scholars of classical literature the most useful functionality to provide, and the only one capable of giving readers of the Digital Library a richer and meaningful experience, is without a doubt reference linking. As soon as a significant number of digital libraries providing a CTS-compliant interface are available over the web, it will be possible to build a client-side application giving readers the capability to move directly from a text reference to its source, and to read or compare different critical editions of the same texts.
Indeed data encoded with the proposed Microformat vocabulary set could be easily mapped into CTS-compliant requests that permit obtaining from a CTS repository - providing that someone would be available – an XML-encoded response containing the requested text passage. However, before it would be possible to build such advanced services, the stability and persistency of existing digital libraries sharing their raw data as a canonical text service needs to be improved considerably.

4.5 A working example

Finally a working example has been supplied to give an idea of the possible uses of microformatted references. The aim of this example is to show that the reference linking system between primary and secondary sources obtained by embedding chunks of information within POSH tags is open-ended (i.e. other resources are allowed to be added afterwards without requiring any modification of the system components), cross-language and flexible.
In this example I examined the possibility of linking canonical text references to relevant resources available over the web by using both microformatted references and "semantic tags". By the term "semantic tag" I mean using semantic information (e.g. URNs) as tags, instead of simple words that just produces language-dependent tags. A similar approach was used in the Perseus Digital Library experimenting with the use of CTS-URNs within queries in Google Base[15] in order to increase the value of its data[Weaver 2007]. Indeed, for each logical section of a work, an item containing its metadata description and the correspondent CTS URN was uploaded to Google Base.
Therefore, I semantically tagged two different kinds of secondary sources on classics that scholars are used to dealing with: some bibliographic records and a popular Italian review of resources, using an on-line service of social bookmarking called Del.icio.us,[16] and CiteUlike,[17] a social platform intended to organize and share bibliographic citations. The review is entitled Rassegna degli Strumenti Informatici per lo Studio dell'Antichità Classica, edited by Alessandro Cristofori[18] and it is a repertory of secondary sources on classics available over the web and is nothing else but a pre-

15 http://base.google.com
16 http://del.icio.us
17 http://www.citeulike.org
18 http://www.rassegna.unibo.it

cise and continuously updated collection of annotated web addresses of electronic resources concerned with classical literature, and often with canonical texts.
Furthermore other kinds of secondary sources that could be tagged taking advantage of social and semantic tagging, include book reviews and iconographic images, because often they are directly concerned with canonical authors, works or text passages.
In order to make the browser aware of microformatted data within the displayed page I used Operator,[19] a Javascript extension written by Michael Kaply for the popular open-source browser Mozilla Firefox, that will be fully integrated within the next browser version. This extension basically works by parsing the entire page and extracting information from the Microformats recognized depending on the Javascript definition of each Microformat itself (provided that it is known by the browser extension). Then I myself extended Operator just adding a few lines of code in order to make the browser aware of the references marked up according to the proposed Microformat vocabulary set.
The result obtained is that the browser understands the meaning of canonical text references displayed inside the page: in fact, it recognizes the author, the work or the textual passage cited. Besides some actions that could be performed upon a given canonical reference are suggested by the Operator extension to the user. Some possible actions are finding relevant bookmarks on Del.icio.us and searching for pertinent bibliographic entries on CiteULike (the only actions supported by this working example), where by the terms "relevant" and "pertinent" I mean related to semantic information expressed by the encoded reference.
For example, the review, written by Gregory Nagy which appeared on Bryn Mawr Classical Review,[20] of the first volume of West's edition of the Homeric Iliad (containing books I-XII) is a resource relating to references such as Hom. *Il.* 1.1, Hom. *Il.* XI or Δ 4. Since this resource was tagged with the appropriate CTS URNs as tags, it now becomes possible to aggregate and provide the user with some information (such as research articles, book reviews, blogpost, conference announcements etc.) concerned with references to books 1-12 of Homer's Iliad contained within the web page.
Given that this example was provided just to instances the capabilities of such a technique, the number of relevant resources would increase as much as scholars using services of social bookmarking do use such "semantic tags" when they tag interesting resources, related to a canonical author, work or text passage.

5. Conclusion

In conclusion the approach described in this paper produces as its result the clear benefit that the semantic meaning of microformatted canonical text references is expressed in a language-neutral, fully semantic and reusable way. Thus it becomes

19 http://www.kaply.com/weblog/operator
20 http://ccat.sas.upenn.edu/bmcr/2000/2000-09-12.html

possible to reach a coherent linking system for primary and secondary sources independently of solutions peculiar to certain applications and therefore not interconnected with each other.
Furthermore, when on-line secondary sources are published on the web with a touch of semantic information in addition and they will provide as a value added service a similar system aimed at linking together primary and secondary sources, therefore scholars themselves will probably be more likely to use the Digital Library for their discipline-specific purposes.

References

Allsopp, John. 2007. *Microformats: empowering your markup for Web 2.0*. S.l.: Friends of ED.

Crane, Gregory — David Bamman — Lisa Cerrato — Alison Jones — David M. Mimno — Adrian Packel — David Sculley and Gabriel Weaver. 2006. Beyond Digital Incunabula: Modeling the Next Generation of Digital Libraries, In: *Research and Advanced Technology for Digital Libraries. Proceedings of 10th European Conference, ECDL 2006, Alicante, Spain, September 17-22, 2006*, pp. 353-366. http://dx.doi.org/10.1007/11863878_30 (2008.01.03).

Fielding, Roy T. and Richard N. Taylor. 2002. Principled design of the modern Web architecture. *ACM Transactions on Internet Technology*. 2(2):115-150. http://doi.acm.org/10.1145/514183.514185 (2008.01.07).

Graf, Alexander. 2007. RDF A VS. Microformats. Technical Report, DERI - Digital Enterprise Research Institute, DERI Innsbruck, Austria. http://www.sti-innsbruck.at/fileadmin/documents/technical_report/html_metadata/RDFaVsMicroformats.pdf (2008.01.13).

Hausenblas, Michael — Wolfgang Slany and Danny Ayers. 2007. A Performance and Scalability Metric for Virtual RDF Graphs. In: *Proceedings of Scripting for the Semantic Web Workshop at the ESWC*, Innsbruck, Austria, May 30 2007, CEUR Workshop Proceedings, 10 p. http://ceur-ws.org/Vol-248/paper2.pdf (2007.06.12).

Khare, Rohit and Tantek Çelik. 2006. Microformats: a pragmatic path to the semantic web, In: *Proceedings of the 15th international conference on World Wide Web*, ACM, Edinburgh, Scotland, pp. 865-866.

Smith, Neel. 2004. TextServer: Toward a Protocol for Describing Libraries. *Classics@* 2. http://zeus.chsdc.org/chs/issue_2_-_smith (2008.01.03).

Weaver, Gabe. 2007. Adding Value to Open Scholarly Content. *Gabriel Weaver's Personal Blog*. http://blog.gabrielweaver.com/2007/03/test-post.html (2008.01.07).