The Use of the Dublin Core in Web Annotation Programs

D. Grant Campbell
Faculty of Information and Media Studies
University of Western Ontario
gcampbel@uwo.ca

Abstract

This paper examines the implications of annotation programs, such as Annotea, for the development of the Dublin Core. Annotation programs enable multiple users, situated far apart, to comment on a Web-mounted document, even when they lack write access, through the use of annotation servers. Early indications suggest that the Dublin Core can significantly enhance the collaborative authoring process, especially if the full set of elements is used in a project that involves large numbers of users. However, the task of adapting DC elements and qualifiers for use in annotation threatens to increase the complexity of the scheme, and takes the Dublin Core far from its connections to traditional library cataloguing.

Keywords: *Annotation programs; Semantic Web.*

1. Introduction

The Dublin Core is expressly committed to fostering the development of metadata description across multiple domains, and to facilitating the interoperability necessary for cross-domain resource discovery [6]. Its development has been an extended exercise in compromise, consensus-building, and dialogue among many stakeholders, including the library community and the web development community. As a result, the Dublin Core has one foot securely in the traditions of information organization; it provides a means of describing electronic resources in a way that can be mapped to traditional cataloguing standards such as the Anglo-American Cataloguing Rules, as well as frameworks for the interchange of bibliographic data, such as MARC. Its other foot rests in the emerging standards that will form the Web of the future, particularly the developments of the Semantic Web under the auspices of the World Wide Web Consortium: XML and its related standards such as Xpointer, the Resource Description Framework, and ontology creation.

These emerging standards of Web design involve not just resource discovery, but resource creation. Building on the democratizing effects of the existing World Wide Web, which has made widespread information dissemination possible to many who are shut out from the traditional publishing process, the Semantic Web seeks to broaden the Web still further by facilitating the creative process of authoring itself. Annotation programs, such as the W3C's *Annotea*, which is implemented in the W3C's Amaya browser, enable multiple users to annotate an existing document without having write access to the document's original page.

The Dublin Core stands ready to play a significant role in these annotation programs as they develop. The nature of this role, however, depends on how ambitiously the DC elements are used. And if used to its full potential in annotation, the Dublin Core could make a significant break from the document-centered cataloguing traditions that played an important part in its development.

2. Annotation Programs

The practice of annotation—providing commentary on information objects created at other times and usually by other people—is emerging as an important dimension of current efforts to facilitate the access and use of information on the World Wide Web. Annotation finds its most obvious use in multimedia systems, where images, sound and video can be annotated with text to facilitate retrieval. Current programs in this area range from simple captioning systems [2] to ambitious and sophisticated systems that provide multiple views of annotations in multiple formats [12].

Annotation also facilitates information retrieval in general, providing retrieval systems with additional means of eliminating spurious matches, and allowing for communication between different users of the same document store [3, 5, 9]. They also have uses in

knowledge management, by enabling organizations to tap the unofficial knowledge base of its members [10], as well as facilitating information evaluation [16].

Annotation services have always played an important part of Tim Berners-Lee's vision of a collaborative and creative Web environment:

> Imagine having servers for comments in different forums, perhaps family, school, and company. Each point and rebuttal is linked, so everyone can see at a glance the direct agreements and contradictions and the supporting evidence for each view, such that anything could be contested by the people involved. … Again, the theme is human beings doing the thinking and machines helping it work on a larger scale, but nothing replacing wisdom in the end. [1]

Although the Web has been slower than Berners-Lee hoped at developing authoring tools, the interest in annotation programs to facilitate collaborative work is growing, as programs like Annotate!, Virtual Notes and DCRS experiment with the process of making user comments available to communities for purposes of collaborative web authoring [14, 15, 17]. The World Wide Web's contribution to this area is *Annotea*, a program that enables multiple users to provide metadata for a single pool of documents for purposes of collaborative writing and research. Three levels of use are envisioned:

- A basic level, at which annotations are used to provide commentary on a single set of documents, according to a set of categories that can be home-grown or standardized;
- A A higher level, at which both resources and annotations are bookmarked according to home-grown

or standardized categories, to generate a variety of resources and metadata displays; and
- A An advanced level, at which the user-provided annotations are supplemented by metadata from other ontologies, often automatically generated. [13]

The default settings for an annotation in the W3C's Amaya browser assigns the annotation values for Title, Author, Source document, the annotation type, the date created and the date last modified (See Figure 1).

Other annotation programs, such as CREAM (CREAting Metadata), are more closely geared to the ultimate objectives of the Semantic Web, enabling either the author or another user to annotate data elements within a document with RDF metadata. Such metadata describes the data elements according to an external ontology, and clarifies their relationship with other data elements, thereby facilitating the document's use by intelligent agents [11].

Whether the task involves collaboration on the creation of a Web resource, or using an agent to assemble virtual documents in response to a specific query, the challenges facing annotation programs are formidable. Once the annotation project grows past a very few users, problems of interoperability, identification, security and timeliness present themselves. The program must be able to provide each user with the most recent annotations, and to assemble annotations efficiently, from each class of annotation, especially when classes specifically tailored to the project at hand have been created. Access and annotation rights must be limited to those authorized at each stage of the process, to preserve confidentiality.

3. The Dublin Core in Annotation Programs

Because of the need for interoperability, identification and access rights, the Dublin Core has a useful role to play for the annotation process. Certainly, the Dublin Core arose partly out of the recognition that metadata needed to be added at the document creation stage, and that widespread acceptance of the Core would encourage software designers to facilitate easy entry by authors [4].

Koivunen and Swick envision the Dublin Core being used to standardize the basic metadata of the annotation. Elements such as the title of the annotation, the name of the annotator, and the date created could be specified as Dublin Core elements, while other elements more specific to the annotation process could either create or use another scheme (see Figure 2).

Other programs, such as CREAM, resist the use of the Dublin Core, on the argument that the metadata, if it is to be used to facilitate the advanced semantic activity envisioned by the makers of the Semantic
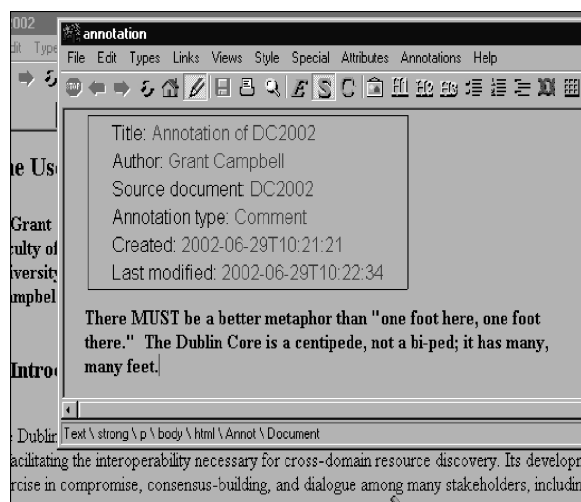

**Figure 1. Annotation using
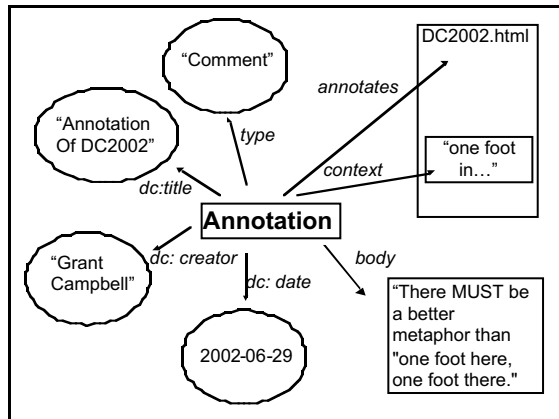the Amaya Web Browser**

**Figure 2. Typical Annotation Scheme
(Adapted from Koivunen and Swick, 2001)**

Web, must be relational. Rather than generating static notes or commentary, the metadata should be making explicit statements of relationships between class instances. According to this argument, annotation programs need to provide more than templates for entering comments, and instead provide identifiers that enable semantic relationships [11].

4. Enhanced Use of the Dublin Core in Annotea

The Dublin Core, however, has more relevance to annotation and collaborative creativity than either of these arguments suggest. In particular, it is possible to use DC elements more widely in the annotation process. Apart from the standard elements of Title, Creator, Contributor and Date, it is conceivable that other DC Elements could be used to express important information about the annotation for purposes of future harvesting, collocation and display.

Some DC Elements could be useful when annotation extends to a wide range of collaborators. These include:

- A Language: for use with documents being annotated across linguistic boundaries.
- A Format: for controlling the styling and display of annotations in different formats, such as HTML and XML.
- A Publisher: for annotation projects that involve a variety of individuals from multiple organizations, this element could be used to link commentators to their parent institutions.
- A Identifier: for providing the URI of the annotation.

Other elements could be used for the actual content of the annotation, as well as some of its important related information.

4.1 Description

The Dublin Core Reference Description defines this element as "an account of the content of the resource" [7]. Typically, it is used for abstracts, tables of contents, or some other graphical or free-text account. The text of the annotation could easily be placed in the Description element. However, such a practice does introduce an element of confusion, since the annotation functions as metadata for the original page, while the Description element serves as metadata for the annotation. Furthermore, the term "Description" does not completely apply to the spirit and purpose of annotation, which is comprises such activities as commentary, criticism, expansion, querying and references to other, related resources.

4.2 Type

Annotea provides a default list of annotation types, such as "advice", "change", "comment" or "question" (see Figure 3). As a description of "the nature or genre of the content of the resource" [7], the Type element could be used to classify the annotation according to a working list of categories established by the group. This would be advisable if multiple documents were being created by various subgroups that would later need to be joined together.

The question then arises: to what degree should the Dublin Core provide qualifiers to the "Type" element to facilitate annotation activities? Certain activities, such as comment, change and question might be considered universal, and worth defining at the level of the metadata set for interoperability purposes. Others may well be defined by a specific group for its own purposes.

4.3 Coverage

While coverage is usually conceived in temporal or geographical terms, it could also be used in a collaborative context to indicate:

- The range of annotation. In this way, aggressive and far-reaching commentary, appropriate to the initial brainstorming stages of a project, could be separated from the grammatical, stylistic and technical annotations appropriate for the proof-reading stages.
- The area of the document covered; annotations of one section, such as the introduction could then be separated from those directed at others, such as the bibliography, or FAQ page.
- The stage of consultation: annotations on an annual report, for instance, could be classed according to those provided by the original team of authors, those provided by the organization as a whole, those provided by government or other external officials, and those provided by the general public.
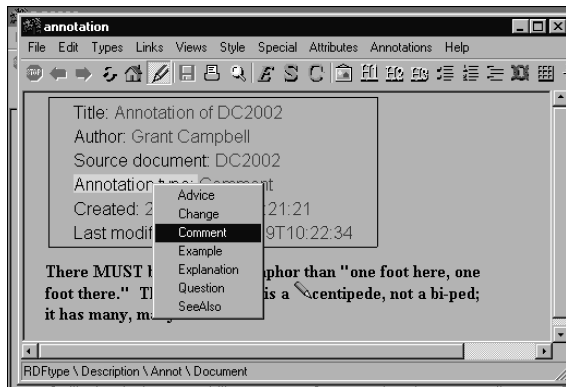
**Figure 3. Annotation Types
in the Amaya Browser**

**Figure 4. Using the Relation Element
to Express a Relationship**

4.4 Rights

This element could be used to administer access to the various annotations. In the case of a document on a sensitive subject which reflected the collective thought of a committee or other administrative body, such administration would preserve the privacy of those involved in the original deliberations, after the point when the committee's decision is made.

4.5 Relation

This element references "a related resource" [7], and as such may be the most significant element for the Dublin Core as it adapts both to collaborative authoring and to the Semantic Web. At present, the established DC refinements support such relationships as versioning, replacement, and part relations. In a simple annotation process, this element could be used for the URI of the original document. With an expanded list of refinements, this element could also be used for relational metadata, thereby embedding semantic relationships that could be used for sophisticated machine processing. In the Figure 4, for instance, an annotation of the author's name creates a link to the author's home faculty. Such a link helps to identify the author as the "Grant Campbell" who teaches for the Faculty of Information and Media Studies, and disambiguates him from others with the same name.

By using the Dublin Core elements to their full potential, therefore, an annotation could look something like this:

```
<?xml version = "1.0"?>
<RDF
xmlns = "http://www.w3.org/TR/1999/REC-rdf-syntax-19990222#"
xmlns:DC = "http://metadata.net/dstc/DC-10-EN/#">
<Description xml:lang="en">
<DC:Title>Annotation of DC2002</DC:Title>
<DC:Creator>Campbell, Grant</DC:Creator>
```
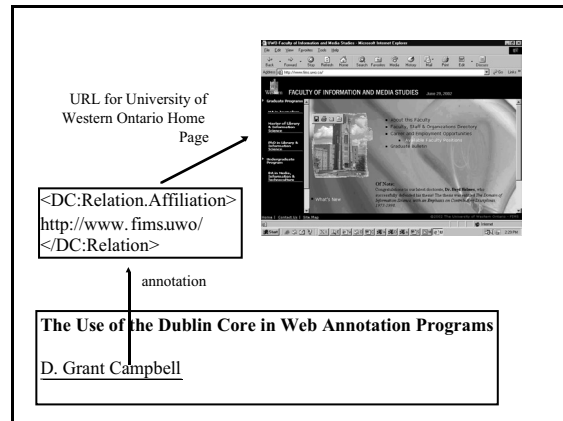```
<DC:Description>
There MUST be a better metaphor than "one foot here,
one foot there."
</DC:Description>
<DC:Publisher> Faculty of Information and Media
Studies, University of Western Ontario
</DC:Publisher>
<DC:Date DC:Scheme="ISO8601">2002-06-29T10:
21:21</DC:Date>
<DC:Type>Commentary</DC:Type>
<DC:Format DC:Scheme="IMT">HTML </DC:Format>
<DC:Identifier DC:Scheme="URI"> http://instruct.uwo.ca/
fim-lis/502/ </DC:Identifier>
<DC:Language   DC:Scheme="RFC1766">   EN</
DC:Language>
<DC:Relation.Annotates>http://instruct.uwo.ca/fim-lis/
502/dc2002.htm</DC:Relation>
<DC:Rights>For use within the authoring circle
only.</DC:Rights>
</Description>
</RDF>
```

## 5. Consequences and Conclusions

Widespread use of the Dublin Core in annotation programs could be highly beneficial to the annotation process. Use of DC elements and qualifiers could simplify the collation and ordering of annotations through standardized versions of dates and formats, and many of the elements could be used, as shown, to do justice to the rich complexity of the collaborative process involved in conceiving, creating, revising and disseminating documents. The annotation process, however, revives certain tensions that have always plagued traditional methods of document description and retrieval, while simultaneously breaking down many of the distinctions that have given these methods their coherence. Making annotation a recognized and important part of the Dublin

Core's purpose could carry profound consequences for DC activities and developments.

### 5.1 Detail vs. Speed

The use of DC elements for extensive support of annotation threatens to revive the structuralist/ minimalist debate that has plagued the Dublin Core for years. While many of the elements can be meaningfully adapted, some, such as the "Relation" and "Type" elements, will need further qualification. And annotation users will be sorely tempted to "smarten up" the Dublin Core to do justice to the subtleties and rich demands of the collaborative process, just as the cataloguing community has introduced qualifiers and refinements to enhance interoperability with MARC records. If annotation programs continue to proliferate, developers may well decide to limit use of the Dublin Core to the few elements that can be simply and unambiguously applied, choosing to extend it with new schemas and alternate schemes as desired. Extensibility, after all, is a fundamental principle of a metadata set that strives to be a core, not a comprehensive descriptive code [8].

### 5.2 Document vs Data

Beneath this revival of the structuralist/ minimalist controversy lies an even more interesting trend. With annotation programs, the Dublin Core is finally moving into a new bibliographic universe: one that we've always been aware of, but have been only fitfully able to inhabit. Unlike the traditional bibliographic universe, which consists of physical documents which are aggregated by catalogues into meaningful units such as editions, series and works, this new universe is highly granular, and breaks documents down before aggregating the individual data elements. This universe, which owes as much to computer science and database design as to traditional library science, is a universe of "entities" and "relationships": terms which are deliberately amorphous and vague, and whose meanings are assigned locally within specific communities and domains, linked tenuously together by ontologies.

This universe of data given local context and assembled across domains through ontologies is, of course, the universe of the Semantic Web. And annotation programs in their various uses break down distinctions that have traditionally prevailed in document organization and description. Annotation reduces the gulf between textual and non-textual information sources by providing a textual dimension to multimedia artifacts. It breaks down the distinction between official and non-official publication, by facilitating unofficial comment on official documents, thereby mobilizing the vast amount of hidden knowledge available in a community or a workforce. And annotation collapses the distinction between information retrieval and information evaluation, by

bringing the community into the retrieval process, and providing additional means by which information can be evaluated at the retrieval stage.

As the Dublin Core moves towards the envisioned world of the Semantic Web, it stands to benefit from the foresight of its initial founders, who, in 1995, chose to address the problem of describing digital objects in general, rather than specifying electronic "documents", "books", or "articles". With the rise of annotation programs, we can see the movement of the Dublin Core away from the "document", whether it is a resource in CORC that must be represented either in DC or MARC, or a resource harvested through an OAI harvesting system. The Dublin Core is breaking through that document layer, and is now describing and addressing discrete data units that can be detached, collated and assembled in fresh and dynamic ways.

## References

[1] Berners-Lee, T. and Fischetti, M. *Weaving the Web: the Original Design and Ultimate Destiny of the World Wide Web*. New York: HarperBusiness, 2000.

[2] Bertino, E., B. Catania, and G.P. Zarri. "A Conceptual Annotation Approach to Indexing in a Web-Based Information System". *International Conference on Advance Issues of E-Commerce and Web-Based Information Systems*. WECWIS, 1999, pp. 160-165.

[3] Bremer, J.M. and M. Gertz. "Web Data Indexing through Semantic-Carrying Annotations". *Proceedings of the 11th Annual Workshop on Research Issues in Data Engineering*, 2001, pp. 69-76.

[4] Caplan, P. and R. Guenther. "Metadata for Internet Resources: The Dublin Core Metadata Elements Set and its Mapping to USMARC". *Cataloging & Classification Quarterly*, vol. 22, pp. 43-58, 1996.

[5] Czejdo, B., J. Dinsmore, C.H. Hwang, R. Miller, and M. Rusinkiewicz. "Automatic Generation of Ontology Based Annotations in XML and their Use in Retrieval Systems". *Proceedings of the First International Conference on Web Information Systems Engineering*, 2000, pp. 296-300.

[6] Dekkers, M. and S. Weibel, S. "Dublin Core Metadata Initiative progress report and workplan for 2002". [online] *D-Lib Magazine*, 8(3). Available from: http://www.dlib.org/dlib/february02/weibel/02weibel.html [June 29, 2002].

[7] Dublin Core Metadata Initiative, *Dublin Core Metadata Element Set*, Version 1.1: Reference description. [online] Available from:

http://dublincore.org/documents/dces/ [June 29, 2002].

[8] Duval, E., W. Hodgin, S. Sutton, and S. Weibel, "Metadata principles and practicalities". [online] *D-Lib Magazine*, 8(4). Available from: http://www.dlib.org/dlib/april02/weibel/04/ weibel.html [June 29, 2002].

[9] Ginsburg, M. "Annotate!: A Tool for Collaborative Information Retrieval". *Proceedings of the Seventh IEEE International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises*, 1998, pp. 75-80.

[10] Ginsburg, M. and A. Kambil. "Annotate: A Web-based Knowledge Management Support System". *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences*, 1999.

[11] Handsuch, S. and S. Staab, "Authoring and annotation of Web pages in CREAM." [online] In *Proceedings of the 11th International World Wide Web Conference, Honolulu, May 7-11, 2002*. ACM. Available from: http://www2002.org/CDROM/refereed/506/index.html [June 29, 2002].

[12] Hirotsu, T., T. Takada, S. Aoyagi, K. Sato and T. Sugawara. "Cmew/U: A Multimedia Web Annotation Sharing System", TENCON 99: *Proceedings of the IEEE Region 10 Conference*. Vol 1, 1999, pp. 356-359.

[13] Koivunen, M. and R. Swick. "Metadata based annotation infrastructure offers flexibility and extensibility for collaborative applications and beyond." [online] In *K-CAP 2001: First International Conference on Knowledge Capture, Victoria 21-23 October*. Available from: http://sern.ucalgary.ca/ksi/K-CAP/K-CAP2001/ [June 29, 2002].

[14] Koch, S. and G. Schneider. "Implementation of an Annotation Service on the WWW: Virtual Notes." *Proceedings of the 8th Euromicro Workshop on Parallel and Distributed Processing*, 1999, pp. 92-98.

[15] Richy, H. and G. Lorette. "On-line Correction of Web Pages." *Proceedings of the 5th International Conference on Document Analysis and Recognition*, 1999, pp. 581-584.

[16] Sannomiva, T., T. Amagasa, M. Yokishawa, and S. Uemura. "A Framework for Sharing Personal Annotations on Web Resources Using XML." *Proceedings of the 2001 Workshop on Information Technology and Virtual Enterprises*, 2001, pp. 40-48.

[17] Tsai, S.R., J. Chen, and M. Koo. "A Document Workspace for Collaboration and Annotation Based on XML Technology", *Proceedings of the 2000 International Symposium on Multimedia Software Engineering*, 2000, pp. 165-172.