

Richard Light  
Independent Consultant / Treasurer, CIDOC  
3 Midfields Walk  
Burgess Hill  
West Sussex  
RH15 8JA  
United Kingdom  
Phone: +44 1444 232067  
richard@light.demon.co.uk

### **The CIDOC web site – a case study in information re-use**

#### **Summary**

The current CIDOC web site has become increasingly difficult to maintain. It relies on the willingness of one member to update its content, and on other members providing him with relevant material. Although it provides French as well as English content, it does so by including both languages in a single HTML page. Content from the web site cannot easily be used for other purposes, such as printed publications.

This paper describes current efforts to re-invent the CIDOC web site so that it is an effective information resource for its users – the membership of CIDOC. The content of individual pages is now held as XML, allowing it to be used for other purposes. French and English versions of each page are stored separately, making it easier to work in a single language of one's choice and opening up the possibility of adding other languages in the future. The whole "site map" is stored as a Topic Map, allowing a range of approaches to navigation to be deployed, including CRM-based filtering.

An online site updating framework has been developed. This allows suitably authorised CIDOC members to add and update both the structure and the content of the site directly, without the need for an intermediary.

During the coming months we intend to issue newsletter articles on the web site, as PDF documents, and in print. The challenges encountered and lessons learnt while doing so will be described.

#### **Setting the scene: why change the web site?**

The current CIDOC web site has become increasingly difficult to maintain. It relies on the goodwill of one member, who hosts the site on his own web space and keeps its content up to date. It also relies on other CIDOC members providing him with relevant new and updated material for the site. In practice this does not work very well, and the site is now very out of date.

Although the site provides French as well as English content, it does so by including both languages in a single HTML page. This makes it difficult to

browse the site in your preferred language. Any attempt to add further languages would only increase the complexity of these “multi-lingual” pages.

Some years ago ICOM gave CIDOC a grant to develop a “template” for publications by International Committees. Over time, the perceived need changed from a traditional word processor template to a more powerful framework within which publications could be developed. Either way, something needed to be done about this project.

Content from the web site cannot easily be used for other purposes, such as printed publications. For some time now CIDOC has aimed to issue its newsletter content on the web as well as in print, and this has involved an extra conversion step. Redesigning the web site opened up the possibility that we might publish to multiple platforms in an integrated manner.

Finally, it should be noted that this was a problem which CIDOC had tried to address on two previous occasions. In both cases, the difficulties involved in liaising between a widely-spread Board and a traditional web design company led to a failure to achieve concrete results.

### **A new approach: separate content and presentation**

It was clear that a more radical approach was needed if we were to make progress with this long-standing problem. I suggested to the Board that we could adopt a web publishing framework which separated the content of the site from its presentation as web pages. The Cocoon software was proposed: it is platform-independent, Open Source (from the Apache Software Foundation) and uses XML<sup>i</sup> both for the content itself and for the control structures which define the layout of the site.

The idea of separating content from presentation is a response to the problems encountered in developing and maintaining a web site, where different people are responsible for different aspects of the site. For example, the people who add and update pages should be free to concentrate on the content they are responsible for, without worrying about how their pages will be integrated within the overall site framework. Similarly designers, responsible for the visual appearance of the site and for navigation strategies, should be able to make changes to the logic of the site without having to update every single page.

Another advantage of the Cocoon framework is its use of XML as the standard storage format for page content. Unlike HTML pages, XML documents can easily be converted to a range of different formats using the W3C’s transformation standard XSLT<sup>ii</sup>. In CIDOC’s case, this means that the XML sources can be used not only for the web site, but also for print publication and for distribution of PDF documents by e-mail. XML is also a much more appropriate standard than HTML for the long-term archival storage of information resources. This is, I suggest, an issue on which CIDOC should be seen to be leading the way.

### Navigation and the site map

In designing a framework for controlling the overall design of the site, we wanted to be able to separate the site map from the pages themselves. Once again, XML provided the answer: we created a document in the XTM Topic Map<sup>iii</sup> format to act as our “site map”. Topic Maps are a very generalised method of describing any conceptual framework. It lets you define *concepts* (known as “topics”), *relationships* between those concepts (known as “associations”) and *instances* of those concepts (known as “occurrences”).

For our web site, each “topic” is an entry in the site map, while the “associations” are hierarchical links between these entries. “Occurrences” are links to the actual pages describing a given topic. This information is held independently of the pages themselves, which gives us several advantages.

First, the navigation links for each page do not have to be included within the page’s content, which greatly simplifies site maintenance. Second, the structure of the site can be altered at will, simply by changing the topics and associations within the site Topic Map. Third, it is possible to change the way that site navigation information is presented to the user, simply by altering the XSLT transformation which converts the Topic Map into part of each page on the site.

Because Topic Maps are an open standard, we can use other TM-aware software to view the site map. The example below is an Omnigator<sup>iv</sup> view of one topic from the site map, showing its links to other topics.

*omnigator 007*

The free topic map navigator from Ontopia. Powered by the [Ontopia Knowledge Suite](#).

[Welcome](#) | [Topic Map](#) | [Manage](#) | [Customise](#) | [Filter](#) | [Export](#) | [Merge](#) | [Statistics](#) | [Query](#) | [Reload](#) | [Not indexed](#) | [Feedback](#) | [No schema](#)

## Conferences

**Names** No occurrences available for this topic.

- ◆ **Conferences** - Scope: *en*
- ◆ **Les réunions annuelles** - Scope: *fr*

**Related subjects**

- ◆ **superclass-subclass relationship**
  - [1997 Conference](#) (subclass)
  - [1998 Conference](#) (subclass)
  - [2002 Conference](#) (subclass)
  - [2003 Conference](#) (subclass)
  - [2004 Conference](#) (subclass)
- ◆ **superclass-subclass relationship**

Figure 1: Site map topic as viewed in the Omnigator

Topic Maps also give us a built-in way to support multi-lingual pages. Each “topic” can have zero or more “occurrences” (pages), and each of these can have a “scope”. We simply define the “scope” of each page to be the language in which it is written. This allows us to separate out the different language versions of each page while retaining the link between them that they are really the “same” page, making it easy to support a switch from one language to another. While this facility is primarily used to support a bilingual approach (English and French), the technique can in principle support any number of languages.

One refinement we have introduced is the use of the CRM<sup>v</sup> as a filtering mechanism. This was really done just to demonstrate the possibilities of the CRM and of Topic Maps! First, the CRM has been expressed as another Topic Map. Then, each site “topic” is categorised as an instance of a particular CRM Class. Finally, a “filter bar” is added to each page, allowing you to select only topics which are instances of a particular CRM Class. This process respects the CRM class hierarchy, so if you were to filter by Event you would also see all topics relating to Activity, etc. Filtering by “CRM Entity” (the top-level class) means that you see all topics in the site map, since all CRM Classes are descended from CRM Entity.

### **The content of the web site**

The pages which make up the site are stored as XHTML<sup>vi</sup>. This makes it easy to convert them into HTML. Essentially they are just copied by an XSLT transformation, which changes them from XML to HTML syntax. This HTML fragment then becomes part of the page which is built up by Cocoon.

The fact that pages are stored as XML means that they can also be processed in different ways. In particular, they can be converted to a page-oriented format such as PDF for printing. This capability will require the development of a new XSLT transformation to convert our XHTML to paginated documents in the XSL-FO<sup>vii</sup> format. Cocoon is designed to allow this type of support to be easily added.

### **Managing the site**

While the site as described above will work quite happily if provided with suitable data, it would be unrealistic to expect CIDOC members to maintain it by updating a Topic Map, and uploading XHTML pages to it. Therefore we have spent some time developing a web-based management framework (also using Cocoon) which provides a series of menus for managing the site. This framework is password-protected, so we can keep some control over who is able to work on the site. It uses the same Topic Map basis as the main site, which means *inter alia* that it is bilingual<sup>viii</sup>.

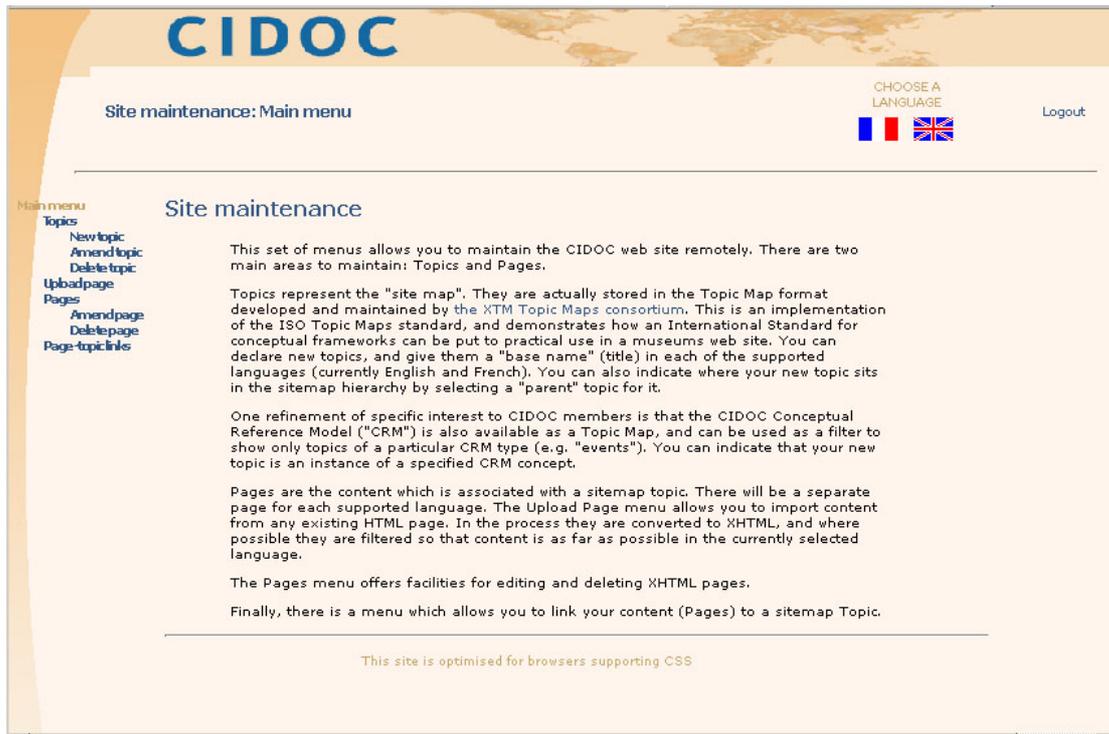


Figure 2: Site maintenance main menu

### Site map maintenance

The main topics menu shows which topics are declared in the CIDOC site. If you select a topic and update the form, its detail will be shown. This detail includes a unique identifier for the topic, the heading it should have in English and French, the topic of which it is a "child", and the CRM Class of which it is an example.

Having selected a topic, you can then create a new topic with this topic as its "parent", or edit it, or delete it from the site map.



Figure 3: Topic maintenance: update an existing topic

One area that needs improvement is that it ought to be possible to control the order in which the “children” of a topic appear. At present they appear in the order they are added to the site map.

#### Content maintenance

New pages are added to the site by uploading an existing HTML page. Cocoon provides a facility to convert HTML to XHTML, and while a page is being uploaded some attempt is made to separate out English or French content, depending on the language currently selected. (This content filtering is aimed particularly at the existing CIDOC site, which includes both languages in each page.) In any case, only the <body> content is transferred, since the HTML in this page will form only a part of the page that is delivered to the new CIDOC site. Before uploading a file, you must specify a site map topic for it to be linked to. During the upload process you can edit the page's content.

Once a page has been uploaded, you can edit it further. There is also a command that lets you delete a page.

### Linking pages to site map topics

Although each page is linked to a site map topic as it is uploaded, you may want to change that assignment in the future. Also, you may want the same information to appear in more than one place on the site. For these reasons there is a menu which shows you the links between site map topics and pages, and allows you to change them.

### CIDOC membership details

A more recent development is an online membership details application. This allows CIDOC members to update and correct their contact details interactively. We hope that this application could ultimately be helpful to ICOM, allowing all ICOM members to update their details via a web form. In the short term it will allow us to keep our own membership details up to date.

### Summary of current situation

Although the bulk of the software development for the new CIDOC site was carried out some time ago, progress with its implementation has been slow. This is mainly because the work still to be done relates to the site map and the content of individual pages.

We discussed this at the Seoul meeting, and worked on the development of an initial outline for the site map. At the time of writing this paper (March 2005) this outline has still to be implemented, but I hope to have done so before the Zagreb meeting. Once this outline is in place, we will need to train volunteers who are willing to add new pages and update existing ones. When this has been done, the new site can be officially launched, and the existing one can finally be retired!

A technical issue which has just come to my attention is that the hosting arrangements for the new site have been disrupted by a change of ownership of the hosting company. It is currently unavailable. While it may be possible to resolve this issue with the current hosts, I feel that for long-term stability we need to investigate moving the site to a more “museum-friendly” host. I will report on progress on this front to the Zagreb meeting.

Finally, some technical developments could usefully be carried out on the underlying framework. As mentioned above, it would be helpful to be able to specify the order in which sub-topics appear. Also, at present we have no web-based facilities for uploading or managing images: these currently have to be dealt with “by hand”. A procedure for outputting pages in “print-ready” format has yet to be developed. Finally, it would be helpful to get feedback from CIDOC members on the overall “look and feel” of the site. One benefit of using the Cocoon framework is that it should be relatively easy to change the look of the site, without having to alter its underlying content.

---

<sup>1</sup> Extensible Markup Language: a World Wide Web Consortium (“W3C”) Recommendation: see <http://www.w3.org/XML/>

- 
- <sup>ii</sup> Another W3C Recommendation: see <http://www.w3.org/Style/XSL/>
- <sup>iii</sup> The underlying Topic Map standard is an ISO Standard: ISO/IEC 13250 Topic Maps (Second Edition) 2002. XTM Topic Maps are an expression of this standard in XML syntax, maintained by TopicsMaps.org: see <http://www.topicmaps.org/xtm/1.0/>
- <sup>iv</sup> The Omnigator is a web application for browsing Topic Maps from Ontopia: see <http://www.ontopia.net/>
- <sup>v</sup> The CIDOC Conceptual Reference Model: see <http://cidoc.ics.forth.gr/index.html>
- <sup>vi</sup> XHTML 1.0 is an XML-based expression of HTML: see <http://www.w3.org/MarkUp/>
- <sup>vii</sup> XSL Formatting Objects: see <http://www.w3.org/Style/XSL/>
- <sup>viii</sup> I would welcome any offers to review and improve my French site maintenance pages!