On the Disappearance of the Library

Torsten Schaßan
Cultural Heritage Computer Science
University of Cologne
Germany

torsten.schassan@uni-koeln.de

Abstract
Digital repositories of various types of material will change the role of the library in the process of information dissemination. The creation of distributed digital resources, which are virtually constructed from holdings existing at different institutions, will lead to an ever increasing interoperability of library resources. Different types of institutional roles in the provision of information for the cultural heritage sector result from this. It may, but need not longer, be combined within one institutional framework, it can also lead to completely new institutional specializations however.

Thesis 1:
According to estimations of the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) within the next 30 to 50 years every single book right now located in a library in print will exist in a digital copy.

Implications:
- A lot of digitization work has to be done. If you consider e.g. the libraries document delivery services, the major problem might not be the number of pages but the rights management.
- If a substantial amount of the sources (literature) relevant to a field of study switch to a new medium, users are willing to change their "default medium of access". This for sure is a two-way street: the more material exists in the new medium, the more users will be using the material and the more users accept the new medium, the more material will be available there.
- Evolving from this are "digital libraries" that differ greatly in structure and use from the libraries today.

Thesis 2:
Libraries today are providers of two semi compatible services: Providers of content and providers of access systems. These services can become independent.

Implications:
- Providers of content. "Libraries" specializing in that service, can provide the largest possible number of "autonomous objects", being prepared for smooth integration with other objects.
- Providers of access systems. "Libraries" specializing in this one, can provide "portals" or whatsoever succeeds them to provide an information architecture for a specific section of the public.

Thesis 3:
Objects in Digital Libraries, if they are "really digitally", tend to behave autonomous. They should be autonomous to accomplish the greatest surplus of the digital medium.

Implications:
- Connections between digital objects will be established ignoring traditional access systems.
- In the same way users of libraries normally are interested in books, not in the libraries themselves. Thus, users want to be able to access digital objects directly, without having to enter the library "through the front door".
- The library will "vanish" from the physical portal and access system to providers of these autonomous objects.
- There might be a development from digital libraries to collections of digital books.

Example 1: The "Digital Autonomous Cultural Object" (DACO)

Imagine the existence of an article on e.g. a medieval manuscript. The article exists in digital form somewhere in the Internet. In this article the author might want to "quote" from a certain manuscript page. By chance, the manuscript page that our author wants to refer to has been digitized and is available in the net as well.

How can the author of the article cite from this manuscript page? The references should follow a system that is in use in the field of research to which the object belongs to, something we use to call "canonical reference". And the references have to be "easy" and self-explaining so that even the not-so-computer-literate historian is able to use them in his work.

And: How can he be sure that his article will not be outdated within shortest period of time because the references are not valid anymore for the library has changed the way of access to the item? To maintain guaranteed access to digital objects is one of the main goals of digital cultural heritage.

Our proposal to answer both questions is, to convert individual cultural heritage objects - books, manuscripts, but conceptually also castles or pieces of furniture - into objects which follow a common behavioural code. That behavioural code provides for the objects informing about them, when asked to do so by a request following the OAI approach, but it also involves the object wrapping itself in mark-up suitable for display in a browser, and integrating itself into a WWW interface, if called correctly. Objects following such a code of behaviour we call *Digital Autonomous Cultural Objects* (DACO).

The behavioural code for a digital object to function digitally autonomous has to provide three things:
- The object has to know *where* it is, to which institution or collection it belongs. Therefore a persistent addressing scheme for collections is needed. Such a scheme necessarily must be organised nationally, with national (or regional) solutions being coordinated by appropriate international bodies.
- The object has to know *who* it is to allow for addressing within the collection. The addressing scheme within individual collections has to be under the control of the individual institution, which guarantees a common functionality and interoperability of the different collections.
- The object has to know *what* it is in terms of granularity. A mapping scheme would allow referencing a granule of a digital object by a specific numbering scheme, which is then translated into the actual names of individual digital components, like page images. Such a mapping scheme is administered by the individual collection and should even exist if the names of the digital objects, e.g. file names, also reflect the traditional references directly.

An implementation of an addressing scheme for digital objects based on the preceding analysis would look as following:

<collection-reference> <object-reference> <granule-reference>

Within one of the digital libraries run by the department Historisch-kulturwissenschaftliche Informationsverarbeitung (roughly: Cultural Heritage Computer Science) at the University of Cologne, the "Codices Electronici Ecclesiae Coloniensis" (CEEC), a complete Cologne codex can currently be reached via a WWW address like:

http://www.ceec.uni-koeln.de/ceec-cgi/kleioc/0010KlCEEC/exec/katk/%22kn28-0083ii%22

The segmentation of the address line according to the above mentioned scheme looks like the following:

<collection-reference>	=	http://www.ceec.uni-koeln.de
<object-reference>	=	ceec-cgi/kleioc/0010KlCEEC/exec/katk/%22kn28-0083ii%22

To access an individual page of a Cologne codex, a WWW address like the following can be used:

http://www.ceec.uni-koeln.de/ceec-cgi/kleioc/0010KlCEEC/exec/pagesma/%22kn28-0083ii_164.jpg%22

Here the following is applicable:

<collection-reference>	=	http://www.ceec.uni-koeln.de
<object-reference>	=	ceec-cgi/kleioc/0010KlCEEC/exec/pagesma/
<granule-reference>	=	%22kn28-0083ii _164.jpg%22
	or	%22|kn28-0083ii _82v.jpg%22

The granule-reference has to distinguish between the direct granule reference and the mapped granule reference.

A direct granule reference consists of a string that can be used directly to access digitised information on a specific server. It may be necessary to break the reference up into components that represent different levels of a storage hierarchy, and/or into components that map logical names unto physical storage addresses. It does not allow for any conceptual interpretation, however. A collection guarantee indicates that the direct granule reference of a digitised page or other atomic unit of digitisation will never change throughout its existence. In our example, kn28-0083ii_164.jpg is a direct-granule reference.

A mapped granule reference consists of a string that is separated by a dividing character, here a vertical line "|". The first of the two parts is the identifier of a mechanism that allows the second part of the string to be mapped to a direct granule reference, according to a specific set of rules, which may be changed over the life span of the digital object or, indeed, be dropped as obsolete. If a mapped granule reference starts with the vertical line, it maps to a default mechanism that will exist for the complete life span of the object and is called a "canonical reference".

Example 2: The "library" as portal to various collections of digital content

At least some of the international research infrastructures have started to connect the results of digitization projects to a decentralized "web", where individual nodes are not "pages", but large repositories. There have been various approaches to this task:

- to proclaim and enforce national standards (Distributed National Electronic Resource (UK))
- to accept many standards / provide standard broker (Research Library Group / Cultural Materials Initiative (US))
- to accept "everything" / provide open broker (Prometheus (Germany))
- to create behavioural protocol for servers for harvester (Open Archives Initiative (international))
- to create behavioural protocol for object servers (DACOs within eCulture-Net)

In the latter case and in analogy to the OAI-approach but less demanding, the servers of collections deliver a list of existing "access venues" encapsulated in HTML or XML, accompanied by RDF / Topic Map explaining "semantic scope". Such access venues might be categories like "people's names" or "content's words". Each of these access venues provides for a list of pages that can be accessed and some expert search elements. The servers deliver functionally complete "objects" to be displayed in a browser. "Functionally completeness" means that the object contains some semantics like e.g. knowledge about previous or next pages, a table of content, etc.

Conclusion: The vanishing of the library

The two examples shown in the paper represent the two types of services that libraries (and any other cultural heritage institution) traditionally offer.

Example one has been the content delivering digital library. Anyway, the digital objects our author refers to from within his article on medieval manuscripts have been delivered to the reader autonomously. The reader does not necessarily recognize the digital library that has provided the digitized page. Yet the library was present all the time – it only has "vanished" into a small icon on the webpage that encapsulated the image.

Example two then has been the access providing system. However the functionality of integration of contents from different databases might be implemented, the assumptions of example one are still valid: the digital objects should be autonomous and thus the library is less visible at this point.

Basic internet addresses:

Example 1:   http://www.ceec.uni-koeln.de/CEEC/texts/Schmitz91.htm
The project CEEC: http://www.ceec.uni-koeln.de
Example 2, ECNet: http://lehre.hki.uni-koeln.de/ECNet