Date : 25/05/2007

Library 2.0 and User-Generated Content
What can the users do for us?

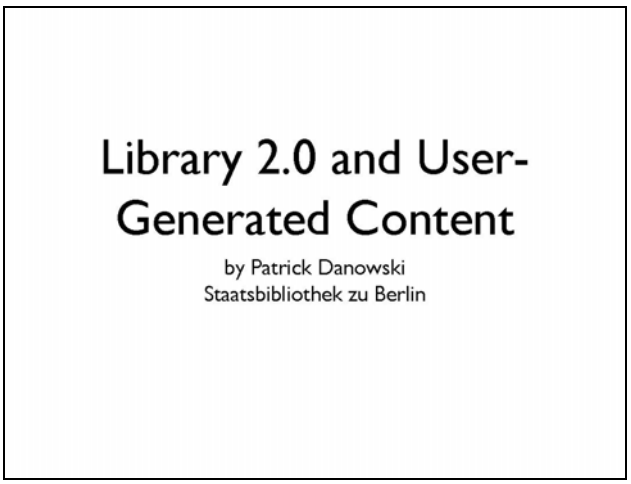
Patrick Danowski
Staatsbibliothek zu Berlin,
Berlin, Germany

Meeting: **113 National Libraries** with **ICABS** and **Information Technology**

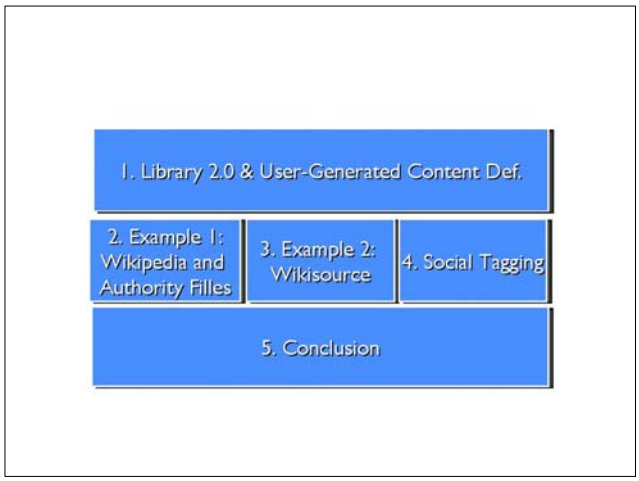
Simultaneous Interpretation: Yes

WORLD LIBRARY AND INFORMATION CONGRESS: 73RD IFLA GENERAL CONFERENCE AND COUNCIL
19-23 August 2007, Durban, South Africa
http://www.ifla.org/iv/ifla73/index.htm

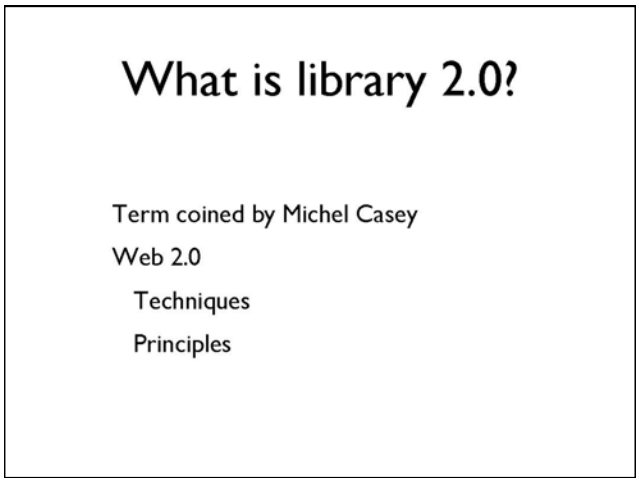
Abstract:
Library 2.0 and user-generated content are two terms, which are closely connected. In the presentation, I will briefly define both terms. Two example projects where user- generated content and libraries interact will be presented. The cooperation of Wikipedia and the Personennamendatei, the German cooperative name authority files is the first. The second will be Wikisource where users provide transcribed source material. Another important area of user-generated content is social tagging where users index different resources. And if the users will do so much in the future, is there still a place for librarians? But in the future user and librarians become partners and the library will provide the platform: the library 2.0.



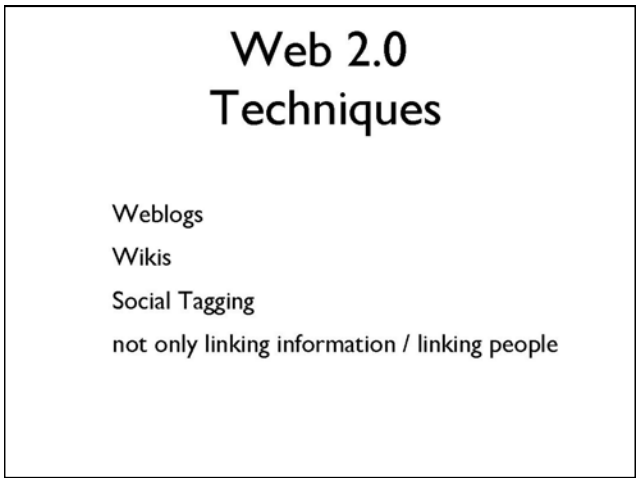
Dear colleagues, I'm very happy to be invited as a young professional. I'm librarian at Staatsbibliothek zu Berlin, Germany's national research library, and in the board of the German Chapter of Wikimedia. I studied computer science and have a master of library and information science. I will not present a concrete project, instead I would like to introduce to you two interesting concepts, that are closely connected.



My presentation will be in five parts. First I want to take a short look at the terms library 2.0 and at user-generated content, what is meant by this? After that I would like to show two example projects, that show that user generated content can be interesting for libraries and how it can be used. My first example is the cooperation between the Deutsche Nationalbibliothek and the German-language Wikipedia project, where the name authority files are used**.** My second example is the project Wikisource. In the German-language Wikisource digitization of books or manuscripts are transcribed by the community while libraries often haven't the money to do that. I will show how the project works and how libraries could create better pre-conditions than there are by now. Certainly we have to take a short look at social tagging. At the end I would like to give a conclusion on what the users can do for us and how we maybe have to rethink a few things. I would like to discuss whether there will be new tasks for libraries in the context of user-generated content.

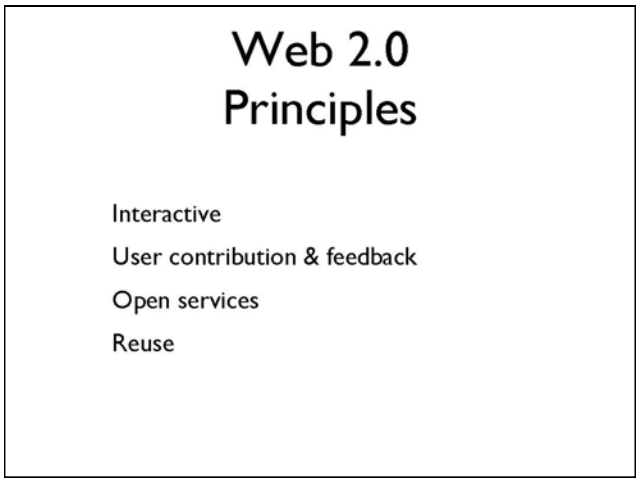


But first I will start with the term "library 2.0". The concept of the “library 2.0” was coined by Michael Casey1. There is not an exact definition of the term “library 2.0” and there have been discussions about it that are still on-going. But common is the point that it borrows from the hype term “web 2.0”2 not only the name. It borrows principles and techniques, too. So let us take a look at the techniques and principles of the web 2.0. Here I just can give you a very small first overview.

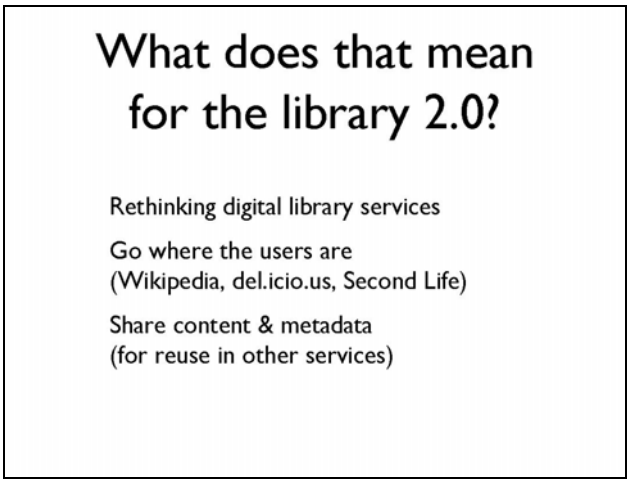


Web 2.0 technology that can be useful in the library can be weblogs3, RSS Feeds4 Podcasts and wikis. All these techniques have in common that users can very fast publish text, sound or in newer services like YouTube videos very easily and quickly. Another type of services very interesting for libraries, is social tagging5 where resources can be indexed by the user through free keywords. First OPACs already are using this technique. Also a very important point of the web 2.0 is that it not only linking information: it is also linking people, one of the most famous sides is MySpace6. Maybe it would be also interesting to use some of these new techniques for “iflanet”7.

1 In his Blog Library Crunch http://www.librarycrunch.com
2 O'Reilly, Tim (2005): What is Web 2.0? Design Patterns and Business Models for the Next Generation of Software. Online: http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html.
3 See the Wikipedia article Blog for a definition: http://en.wikipedia.org/wiki/Blog last visited 28.04.07
4 See the Wikipedia article RSS for a definition: http://en.wikipedia.org/wiki/RSS last visited 28.04.07
5 See the Wikipedia article Collaborative tagging for a definition: http://en.wikipedia.org/wiki/Collaborative tagging last visited 28.04.07
6 MySpace http://www.myspace.com last visited 28.04.07
7 IFLA net http.ifla.net last visited 28.04.07



Principles of web 2.0 that are also applicable to libraries are the interactivity and the possibility for user contribution. Also the feedback from the user should play a role for generating new services and changes of the old ones. The services should be based on open standards. Lawrence Lessig speaks of the Remix-Culture8 and libraries have to become a part of this. Libraries have to make it possible that content (generated by the library) or its services can be remixed in mash-ups.



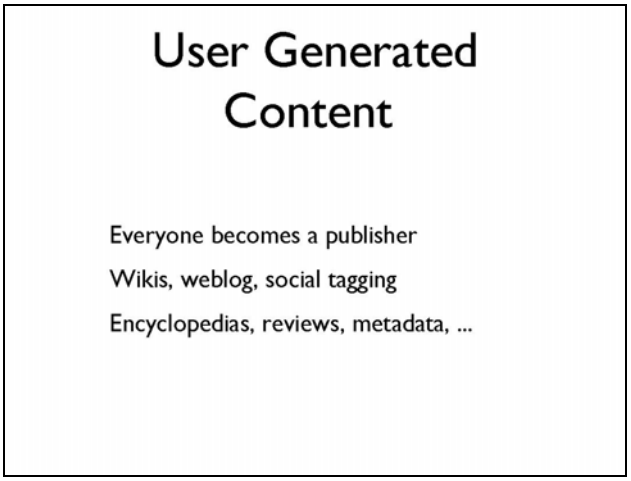
For the library web 2.0 that means that we have to rethink the services of the digital libraries. We don't have to invent everything by ourselves. We can go there, where the users already are. This means that libraries have to consider which web2.0 services are being used by the users and think about how the library can use these services to help the users or bring them closer to the library. Libraries have to create tools (widgets) which can be used in portals like the Google Start Page “iGoogle”9. A second example for that is the use of Second Life10

8 Lessig, Lawrence (2006): Free(ing) Culture for Remix. In: Lutterbeck, Bernd; Gehring, Robert; Bärwolf, Matthias (Hg.): Open Source Jahrbuch 2006. Berlin: Lehmans Media. Online: http://www.opensourcejahrbuch.de/portal/scripts/download?article=osjb2006-09-03-en-lessig.pdf last visited 28.04.07

9 IGoogle Homepage http://www.google.com/ig?hl=en last visited 28.04.07

10 Second Life Homepage http://www.secondlife.com last visited 28.04.07

where American libraries have created Info Island11, a own area in Second Life for libraries and information provider. Libraries can also play a role in services like del.icio.us12. They can import subject guides or lists of links for example and tag them with the keywords they would normally use. A user who works in this subject area will discover this guides more easy cause of the social features and can subscribe to this library resource. Another way to take part is to share their knowledge in Wikipedia. Sharing would be an important point in library 2.0. Sharing content and sharing metadata would be the base for new library services that can face the challenge of the web 2.0. Mash-Up services are also a new trend in the web 2.0 which is very successful.

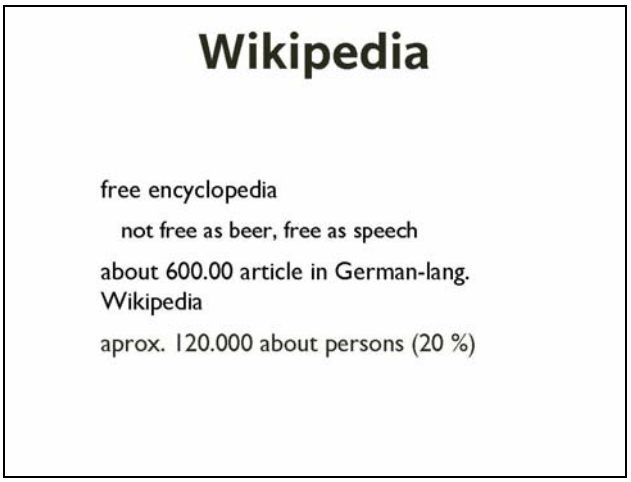


The second term of my title is the user-generated content. The new tools of the web 2.0 like wikis and blogs allow users to become very fast a publisher. Amazon and "library thing" allow users to write reviews for books. There are a lot of new services where users can index different resources like links, photos and videos. But the well known and maybe biggest project of user-generated content is the Wikipedia, the free online encyclopedia which is available in over 100 languages. Wikipedia is free, that means not only the access, or like Richard Stallman calls it “Free like beer”13. Wikipedia is also free for reusing, or in Stallmans words again, its “Free like speech”. Every community can have their own rules, their own project. Since 2005 the German-language Wikipedia has a project together with the Personennamendatei (PND), the German cooperative authority file for personal names. This project is my first example. When I now speak of the Wikipedia, I am referring to the German language project.

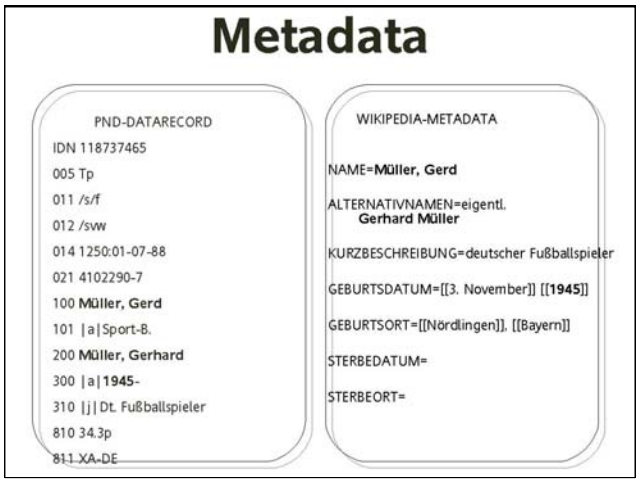
11 Info Island Homepage http://www.infoisland.org/ last visited 28.04.07

12 Heller, Lambert: Libraries that del.cio.us In netbib Weblog http://log.netbib.de/archives/2006/10/05/libraries-that-delicious/ last visited 28.04.07

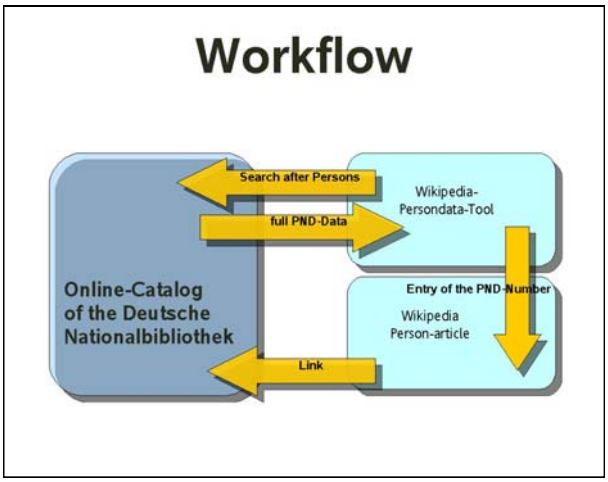
13 Stallman, Richard: „Why "Open Source" Misses the Point of Free Software“. In: Lutterbeck, Bernd; Gehring, Robert; Bärwolf, Matthias (Hg.): Open Source Jahrbuch 2007. Berlin: Lehmans Media. Online: http://www.opensourcejahrbuch.de/portal/scripts/download?article=osjb2007-00-02-en-stallman.pdf



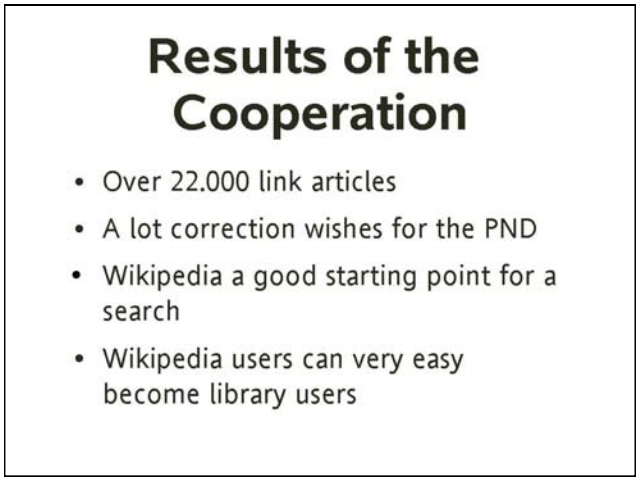
In the German language Wikipedia are over 600.000 articles, from that are about 20 % articles about persons. Metadata for these person articles were created in 2005, too. The reason for that was another project. A publisher produced an offline version of the Wikipedia on CD, but they had a small problem, because they liked to change the order of the names for retrieval purposes. In the Wikipedia usually is the form “FIRSTNAME SURNAME” used. But for the CD they liked to change it in the way that is typically used in encyclopedias:”SURNAME, FIRSTNAME”. The publisher worked together with the community and we created the metadata for person articles.



This was the starting point and some people within the Wikipedia community thought that it would be useful if your data could be connected to the name authority files. Like you see on the chart, we have comparable content: the name, alternative names, the birth and death dates and places and a sort description of the person. The Deutsche Nationalbibliothek agreed to implement a project linking their authority file for personal names with the persons within Wikipedia. They had to do some changes in their database and a developer of the community created a tool to compare the entries of the authority file and the metadata of the Wikipedia.



The workflow was that the Wikipedia Persondata-Tool searched via a special created URL in the OPAC of the German National Library where the authority files are integrated. From the OPAC the full metadata for one or more persons that matched the searched name where returned. The tool presents the metadata of the Wikipedia comparable with the entries of the PND to the user. The user decides whether it were a match and whether the identifier, the PND Number should be added in the article. By means of a template a link was created from the article to the catalog of the German National Library. Following this link you can find all literature from and about this special person in the catalog of the German National Library. This is possible because the search key allowed to retrieve only differentiated personal names.



This cooperation was very successful. In a very short time (about 2 weeks) over 22.000 articles where linked. As reminder, all this linking was made by users of the Wikipedia. During this project a long list of correction wishes of the authority files was created by the community, where members thought there were errors in the authority file. Wikipedia readers can now find further literature for a person very easily. The way from the electronic resource to an offline resource is now very short and a Wikipedia user can become a library user.

Wikisource

Sisterproject of Wikipedia

Choosen Sources:

Interesting and specials sources

which are not aviable in full test

The second example I would like to talk about, is the project Wikisource. Here I will speak about the German language project, too. I will do this because also this project is a little bit special. The goal of the Wikisource project is to provide original source texts. It is clear that these are only texts which have no copyright (Public Domain). The technical basis for this service is also a wiki. With a little extension the proof reading is very easy, because the picture is over the edit window. Libraries can benefit from this service. Users will do what libraries can't afford to do. The result will be certainly free again so that it can be re-used by the library. There is often a discussion about whether libraries should build their own services on their websites, like this one. But this has an crucial disadvantage. Every library has to build its own community, so the critical mass of people for a community that works will be harder to reach. It would be better to allow the reuse of the own material.

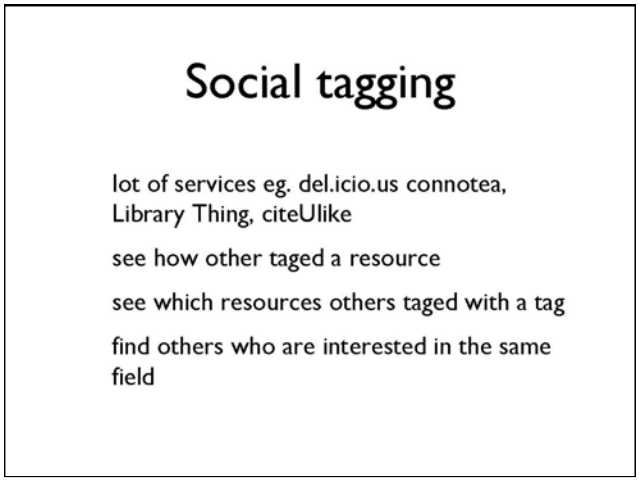
Metadata

- No bibliographic standards
- Goal: cooperation with libraries
- Idea: Downloadable metadata

Not perfect in the moment are the metadata of Wikisource. They are plain text in a table on every cover side. But in the moment there is no way to download only the metadata for using them in other databases for example library catalogs. We already thought about a way to transform the metadata, but we don't have the time to implement this. All people work for this project as volunteers. But even if we did this we still had another problem. The form of entry wouldn't be standardized. We will see how we will solve these problems. If you have any ideas please contact me. About this project I will also present a poster at IFLA and I would like to invite you to talk about this project there.

Like I was explaining in the beginning users are indexing different types of media: links, pictures, videos and text. All these things get tagged.

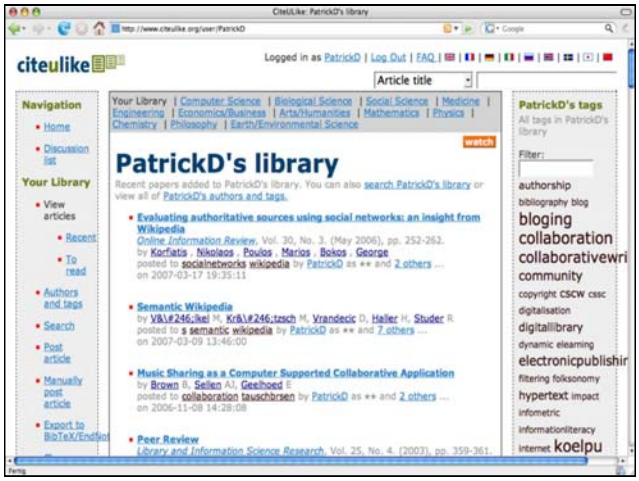
Tags are uncontrolled keywords. So that there can be a lot of tags which mean the same thing or are virtually synonymous. But tags can not only be keywords that describe an object. They can also be functional tags like "impotent" or "have2read" that has only a meaning to this user. Or, a tag like "ifla2007PD" which describes resources of a presentation so it has a meaning for a small group.



More and more services are rising in the web2.0 where tagging is used. The most famous examples are del.icio.us, LibraryThing14, connotea15, citeUlike16, flickr17, and YouTube18. As a user of these services you can certainly tag resources but you can also see how others

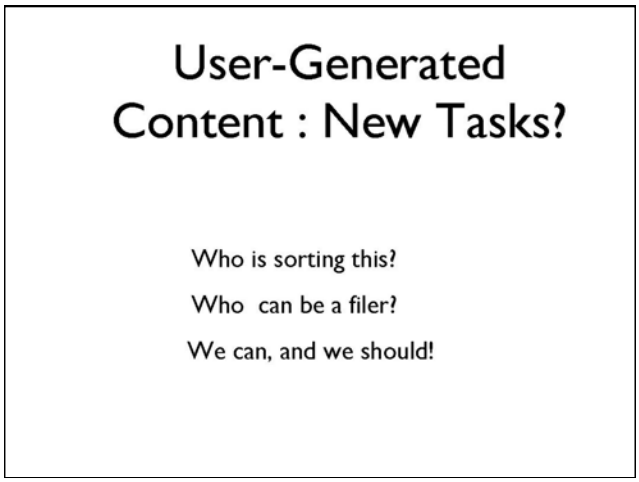
14 Library Thing Homepage http://www.librarything.com
15 Connotea Homepage http://www.connotea.com
16 CiteULike Homape http://www.citeulike.com
17 Flickr Hompage http://www.flickr.com
18 YouTube Homepage http://youtube.com

tagged a resource and who has saved this resource, too. You can search for other resources that have been tagged with a keyword you choose. These will help you to find persons which are interested in the same area like you. That's what I meant at the beginning by saying that the web 2.0 not only links resources but also helps to link people.



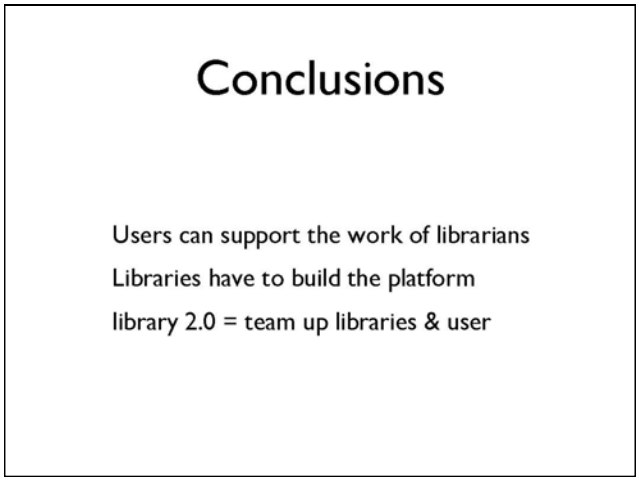
I would like to give a short example for one service. CiteUlike is a service to manage literature, especially journal articles. There are some nice features like you can watch the table of content of some print journals and import the metadata of interesting articles into your account. You can tag every article and you will see how many other people have saved this article, too. I can also enter new articled manual or import a BibTex19 file.

To get a better overview tags can be visualized in a tag cloud. In this clouds tags often used and maybe more relevant would be bigger. Such a cloud is on the right side of the example. The KUG20, the OPAC of the university library of Cologne, uses this technique in their OPAC to visualize the usages of controlled keyword.



19 See the Wikipedia article for more information: http://en.wikipedia.org/wiki/BibTex
20 KUG OPAC http://kug.ub.uni-koeln.de/

We have seen users can do a lot for a library from creating additional information over improve the access to digitalized objects (trough transcription), to index resources. So maybe one question is what should the librarians do in the future? We saw how much different material is on the net. Users can publish it very easy. This means also that there are high quality materials, which can be used in learning environments, and there are materials like the film or photos of the last family vacation. Social Tagging is very powerful but also has its problems. Buzzwords are sometimes used a little bit too often for example. Librarians can create subject guides with the best and the content that can be best used. So librarian can filter and organize the material in their standardized keyword and classification schemes. This classification by the librarian is not the result but the opposite: Librarians' classification can be the basic layer for the web 3.0, the semantic web21. But these systems will get a wider acceptance if they were open and free to use.



Now we have reached our starting point again, and we see that library 2.0 means not only the use of web 2.0 technology; library 2.0 means also that the library becomes an important player in the web 2.0 world. The potential is really big, so get ready, become a librarian 2.0, be open for input from your users. I hope I have shown to you that users can support the work of librarians. We have to create the platform for that, this platform is an important part of the library 2.0. The users can become our new partners, they can help us to improve our services.

I hope I have made you curious because that is always the starting point for something new. Let's try it.

Thank you.

21 See the Wikipedia article for more information: http://en.wikipedia.org/wiki/Semantic_web