

User Benefits and Crowdsourcing – Articles in the Spotlight

Majlis Bremer-Laamanen

The Centre for Preservation and Digitisation, The National Library of Finland,
Mikkeli, Finland

E-mail address: majlis.bremer-laamanen@helsinki.fi



Copyright © 2013 by **Majlis Bremer-Laamanen**. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License:

<http://creativecommons.org/licenses/by/3.0/>

Abstract:

Today we are living in an environment where large user groups want to create and re-create collections for their own needs in collaboration with the communities.

The National Library of Finland will launch one of the first article crowdsourcing projects worldwide in the summer 2013, just in time for the Singapore IFLA Conference. The project was managed by the Scrum-method – an agile software development tool, where the user's point of view is important.

We have created the crowdsourcing tools to enable clipping and adding metadata into the newspaper articles. This enables the users to collect articles in their own interest and to share their interest with others.

The crowdsourcing of articles is embedded into the Finnish Historical Newspaper Library, containing all copyright free newspapers published in Finland from 1771-1910. Later we will also use this crowdsourcing tool for newspapers in copyright. Newspaper houses have already shown an interest in the project and its findings.

What we need is to have a stable user base right from the start: people willing to use the features and also doing it. We hope to engage the large number of current users of the Finnish Historical Newspaper Library and others in community networking.

Keywords: Crowdsourcing, newspapers, articles, The National Library of Finland, The Finnish Historical Newspaper Library.

1 USER BENEFITS AND CROWDSOURCING

Today we are living in an environment where users want to create and re-create collections for their own needs and share them with their communities.

The National Library of Finland is launching one of the first article crowdsourcing sites, just in time for this IFLA Conference in Singapore in 2013.

The NLF has a history since 1640 and it has got legal deposit rights since 1707 to the printed material like books, newspapers, journals and others. Unlike many other libraries the NLF has an inhouse solution for digitisation and preservation, at the Centre for Preservation and Digitisation. The Centre is in charge for the policies and development in these fields.

To this date we have e.g. digitised all the copyright free newspapers, 80% of the Finnish journals until 1944 and industrial ephemera prints. Together this makes 7 million pages. Of these 3 million are in the public domain and thus available for crowdsourcing at our site "Digital Content". This site has almost 8 million yearly page views and contains:

- The Historical newspaper library 1771-1910
- The Historical Journals library 1810-1944
- Industrial Ephemera 1810-1944

2 AIMS AND BUDGET

We are in the middle of the huge task to digitise large corpuses of collections. Our library has sought for a variety of solutions for the funding of this enterprise. But at the same time we have to take into account the future demands in the virtual environment. More diverse search functionalities are asked for by researchers and the public. With shrinking budgets crowdsourcing can be beneficial for the library services and especially for the users themselves.

The task for crowdsourcing has to be interesting for the user. It has to be easy to do and rewarding.

The average user has at least two ways of using the digitised collection:

- from the site in the same way as using the printed version. The image of the print and the bibliographic record is available
- from the content on the site: search from the text enriched with ocr, from marked up articles or images using the METSALTO format. It is here that crowdsourcing can enrich the material.

As the NLF already had an text correction crowdsourcing game going on in 2012 we found it useful/interesting to make a platform for marking articles. The library has an article index for newspapers covering the time period up to 1890 and also dating from that period. However the coverage is under 10 percent. Our users are in need of a better article coverage for all our digitised newspaper and journals collections. The articles could also be interesting for text correction activities in the future.

The project began in late 2011 with a funding of 180.000 e from the European Union Social Fund. The concept has been designed by our staff regarding content and technical matters.

3 OUTSOURCING

During the project, plans were changed. Instead of recruiting staff as we had planned, we got permission from the EU - funder to outsource the manifold IT-design for the application software. At first we prioritised, even more clearly, our aims in the existing concept draft, with the help of an outside partner. Users' input was used and all results were of importance in the bid.

A supply contract was made with the Finnish firms Gofore/Evident in January 2013. We have had a flexible way of working based on the ScrumBan method. Daily scrum 10 minutes sessions with Evident (IT) and Gofore ensured a seamless cooperation and the flexibility we called for. The first version of the Kuvataalkoot, the Alfa version has been launched in late June 2013. It has the features for marking articles and images and simple search functionalities. User feedback during summer and autumn will enable us to come out with the beta version in late 2013.

4 ENRICHING CONTENT – THE USERS

During the project Gofore and the Library made a questionnaire to establish the user needs for article crowdsourcing of the Historical Newspaper Collection. The questionnaire reached out during a two-week period via the Historical Newspaper Site to all interested users. It was also advertised in Facebook - and we got 231 answers of which 20 non-previous users did contribute.

Interest groups for enriching content are:

- the users of the digitised materials: researchers, students, the general public, genealogists – users from all around the world
- the contributors: the reuse and enriching of the digitised materials
- the creators of the original materials

These first two user groups have been in focus in this project. The creators will be addressed when material under copyright will be addressed at a later stage.

The user questionnaire at the Historical newspaper library is guiding our work:

- The typical user is active and uses the digital content often (1-7 times a week) as his/her source. The continuous use supports our aim to get an active user group for the enrichment of the collections.
- Positively, digital content was already shared between users among 22 of the respondents
- The high age of the users correlated with some smaller use of the social media: however 160 users were linked into many social sites, 75 persons did not use social media.

Results of the questionnaire:

- the users want to share their work – collectivity is sought for in the social media and on the site
- the users do want to have a place for privacy, as well, a page with their own remarks
- The site should be easy to use
- Registration of users and sharing of articles should also take place outside the social media. This will in the future enable article crowdsourcing of collections in copyright.

5 BUILDING THE PLATFORM:

When having to choose the most important features in line with the project budget the users' views and feedback were most important. Before that the Digital Content- software had to be updated for the best possible implementation for the Kuvataalkoot-features.

We decided on an architecture including:

- sharing of articles. The articles will be shared automatically, when a user is marking it. As some of the users are not very keen on adding much metadata to the article we will start with:
 - marking of categories as the only compulsory work
 - writing the name of the article is voluntary as are the marking of content groups and content vocabularies (in Finnish and Swedish) .Other words may be added by the user
 - semantic wording like: place, time, event is offered to the user to fill in and comments can be added
 - the name and date and page of the publication is added automatically
- a personal site.
 - Comments of your own and vocabularies can be used
 - the information can be shared with others
- the Kuvataalkoot -site is embedded, interlinked to the Digital Content / Historical newspaper Library site. Here the user finds the articles marked already and can join in.
 - users can join also via Google, Twitter, Facebook
 - users can use the marked articles as a search entrance
 - Using the categories or vocabularies
 - via the publication, name, date

The project has been more beneficial than we thought from the start. The Digital Content software is re-newed to answer the needs of this day. The Kuvataalkoot can be used for all collections on the site and not only for newspapers.

We were happy to see that the users of the questionnaire did give such a positive response to go forward with the project. We have done much prioritisation - more than we would have liked to. As the first priority we have dealt with the PC environment and primary needs of the users. Later on, we will continue, if possible, with mobile devices and semantic, text based and other search possibilities.