



## Crowdsourcing Metrics of Digital Collections

**Tuula Pääkkönen**

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

[tuula.paakkonen@helsinki.fi](mailto:tuula.paakkonen@helsinki.fi)

### Abstract

In the National Library of Finland (NLF) there are millions of digitized newspaper and journal pages, which are openly available via the public website <http://digi.kansalliskirjasto.fi>. To serve users better, last year the front end was completely overhauled with its main aim in crowdsourcing features, e.g., by giving end-users the opportunity to create digital clippings and a personal scrapbook from the digital collections. But how can you know whether crowdsourcing has had an impact? How much crowdsourcing functionalities have been used so far? Did crowdsourcing work? In this paper the statistics and metrics of a recent crowdsourcing effort are analysed across the different digitized material types (newspapers, journals, ephemera). The subjects, categories and keywords given by the users are analysed to see which topics are the most appealing. Some notable public uses of the crowdsourced article clippings are highlighted. These metrics give us indications on how the end-users, based on their own interests, are investigating and using the digital collections. Therefore, the suggested metrics illustrate the versatility of the information needs of the users, varying from citizen science to research purposes. By analysing the user patterns, we can respond to the new needs of the users by making minor changes to accommodate the most active participants, while still making the service more approachable for those who are trying out the functionalities for the first time. Participation in the clippings and annotations can enrich the materials in unexpected ways and can possibly pave the way for opportunities of using crowdsourcing more also in research contexts. This creates more opportunities for the goals of open science since source data

becomes available, making it possible for researchers to reach out to the general public for help. In the long term, utilizing, for example, text mining methods can allow these different end-user segments to achieve more. Based on our current initial experiences, we feel that crowdsourcing gives an opportunity for a library context to get closer to the user base and to obtain insight into the numerous opportunities, which the digitized content provides for them and for the library. Gathering the first prototype qualitative and quantitative metrics for this particular crowdsourcing case gives information on how to further improve both the service and the metrics so that they can give valid information for decision-making.

**Key Words:** digitized collections; crowdsourcing; metrics; monitoring.

## 1. Introduction

The presentation system of the National Library of Finland (NLF) at <http://digi.kansalliskirjasto.fi> contains newspapers, journals and ephemera, currently for a total of over 9 million pages. The presentation layer was recently upgraded and launched on 13.5.2014 to continue to the next level of crowdsourcing, by allowing users to make digital clippings, i.e. cuttings from articles of their choice. The reasons for adding these functionalities were numerous: firstly there was interest in seeing what kind of articles would be considered by the users as most useful for them, and in enabling users to collect their own data set to their own scrapbook. There was also the hypothesis that possibly, when a text mining project goes further than the article extraction, a set of articles created by humans could be used as a comparison set, maybe even as a ground truth, for verifying at least partly the results of the machine learning. Finally, as the clippings could be shared in social media, the aim was to allow users to discuss and compare materials from different sources and possibly to create new interest in the digitized content.

Naturally, now that this new style of crowdsourcing effort has been running for a few months, we wanted to collect metrics of its usage. Did it work? In which contexts were the clippings used? Were the keywords used, and how? What can we learn from our users, based on what they do, and also on what they have told us in their feedback?

This paper is organized as follows. Section two describes the methodology and the motivation behind the metrics. Section three treats the data collection

and the alternative metrics. In section four some of the metrics are analysed in detail, and in the final section we evaluate the quality of the metrics and discuss how these metrics should be developed further.

## 2. Why Measure?

There is a famous quote stating that you cannot manage what you do not measure. The NLF benefits from the legal deposit act of Finland by getting copies of everything that is published in the country. Therefore, it has obligations to both preserve and disseminate this information to everybody, as stated in the NLF vision: "Treasures of the nation to all". Therefore, also in the crowdsourcing efforts initiated by one of the units of the NLF, the Centre for Digitisation and Preservation, the metrics have been a key topic from the start of the crowdsourcing project.

There are a few key metrics in the annual reports sent to all stakeholders and funders (Ministry of Education, library sector, other memory organizations), which are also reported annually. These include, but are not limited to, the number of digitized pages, the number of page downloads and the total amount of the digitized pages (in free or restricted use). The Centre for Digitisation also has internal metrics, which are used to follow website visits and page views, and the crowdsourcing metrics are now a new addition. The metrics are used in following the status of our services with the digital partnerships and the clients, and for the planning of possible new communication efforts.

For a crowdsourcing project the above metrics are not directly applicable. Naturally, crowdsourcing benefits from the additional material digitized and crowdsourcing is one way to engage both existing and new users, and thus to impact the number of page views. Thus, when we launched the crowdsourcing topic, we also had to make new plans for the metrics.

### 2.1. Planning of the Metrics

The pure collected data in the database is just raw material. Metrics help to comprehend and visualize the collected data in a more user-friendly way. This makes the data more usable and gives a better insight into understanding the

user base. The first thing, when thinking of creating the metrics, is to determine which information is actually required. Next, the collecting of the raw data for the metrics is introduced and after following the usage for a while, the first trends can be created from the metrics.

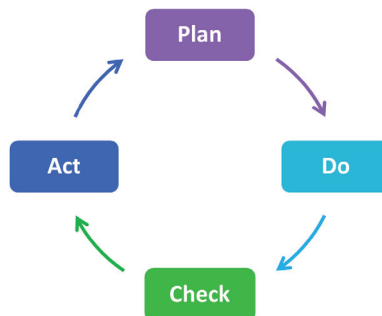
For the process of developing the metrics we used the Plan-Do-Check-Act model, which aims for a systematic process improvement with four simple steps. Despite critique, e.g. by Cole (2002) who claims that the PDCA-model is a bit too rigid for rapidly changing environments, we felt that a light-weight application of the model was suitable, because it gives a structure for the planning of the metrics and is quite straightforward to implement.

Sokovic, Pavletic and Kern (2010) explain the PDCA-model (Figure 1), which was originally developed by Edward W. Deming, by describing the phases:

- Plan: Leadership team defines targets
- Do: Suitable teams start working towards the targets
- Check: Scorecards and metrics are reviewed
- Act: Adjustments after checking have been done and the cycle can start again.

An important part of the PDCA-process improvement is the idea of the development as a cycle. This aspect has important consequences for the usability of the method, and it allows stakeholders to get involved in different phases of the development.

*Fig. 1: Quality circle (Plan-Do-Check-Act model) by Edward W. Deming.*



In the case of the Centre for Preservation and Digitisation (CPD), the discussion surrounding the need of metrics started some time before the launch of the new version of the presentation system of the digitized collections. The initial need was to see how crowdsourcing attracts the interest of the users and what the general consequences of its implementation are. Then, interestingly, the metrics also appeared in the discussions about the project, where copyrighted materials were planned to become accessible for more users via negotiations with copyright organizations and news media companies — both of these parties were interested in the metrics, too. So, in a way, providing metrics could actually be of service both to the library and to the news media companies we collaborate with.

The aim was to develop only a few very basic metrics covering the quantity of clippings made over time: how many keywords are there, and to which materials in the clippings are they targeted to? Another aspect was also to follow the end-user feedback as it exists so that people can directly give input about the features and the contents, or even about extra wishes that they might have.

## **2.2. Do and Check is Actually Develop and Test**

The initial planning gave enough input for developing the metrics. The main purpose was to first get the basic direction, and then to further develop the metrics. This is the process, which is also used in agile software development — first getting a basic understanding about the task at hand with the stakeholders and then evolving until the end result provides the highest possible gain with the minimum effort needed. After initial versions of the reports were developed (Do-phase), they were shown to the stakeholders and then developed further to answer their requirements (Check-phase). The scheduling of the reports was also considered, but at least for the time being the reports work in real-time. It seemed that it might be beneficial to develop the automation further, so that the report can be subscribed by anyone within the unit, who can then send the report to their email.

One of the positive things for the metric development was that a few of the core library metrics had already existed for a number of years, so the basic data collection features were already in place. Therefore, the largest part of the development could focus on the metrics themselves — extracting the data

and developing comparisons and visualizations, which should be both functional and usable.

### **2.3. Actions Onwards**

As typically happens in computer science or software development, when there is an interesting feature, the requirements for the metrics and reporting increase. From the original ideas and needs, a new set of metrics for crowdsourcing is formed. Interestingly, the content of the metrics is now starting to change from the library-oriented metrics towards the content metrics. Crowdsourcing also gives a new boost to the traditional content metrics, and illustrates how different kinds of materials are, and could be used by new types of users.

These new users actually want to participate, and make their own interpretations of the digitized collections provided (Bernstein, 2006). This can be considered as the first step towards the remix culture (Duncum, 2013), where existing works are taken and utilized either by mixing different materials together, or by utilizing existing content in completely different contexts.

## **3. Crowdsourcing Background**

### **3.1. First Crowdsourcing Project — Digitalkoot 1.0**

The National Library of Finland has been enthusiastically experimenting with crowdsourcing activities from early on. In 2011, Digitalkoot 1.0 was launched, and with it the OCR recognition errors were targeted to be fixed with the help of users via a game (also known as the “molegame”) (Bremer-Laamanen, 2014). The system used the typical PBL gamification elements, meaning Points, Badges and Leaderboards. All-in-all, there were nearly 110 000 participants who completed 8 million fixing tasks in total (Digitalkoot, n.d.). Media visibility was also quite considerable, as the news of the crowdsourcing effort went viral and it was picked up by major news sources both globally and nationally by different media, such as The New York Times, Wired, Helsingin Sanomat. Throughout the game process the corrections were stored and even recently the words corrected were used to help in the analysis of the current OCR quality of the newspaper corpus (Kettunen et al., 2014).

### **3.2. New Crowdsourcing Effort — Digitalkoot 2.0**

The new #Digitalkoot crowdsourced project was launched on 13<sup>th</sup> of May 2014. In this case, the aim of the project differed from the earlier project (gamified OCR correction of newspaper text). The idea was to let people use and collect material interesting for them via digital clippings. The clippings could basically be any content from the newspapers, journals or ephemera: for example advertisements, news stories or pictures.

What needs to be highlighted is the fact that making a clipping requires end-users to follow quite a rigorous process with specific steps. In the following section, the clipping workflow is described in order to explain the process, which has led to certain kinds of metrics to be collected.

#### **3.2.1. Search**

Utilizing the digitized collections starts from the user need — what is the topic of interest, what does the person want to investigate, study, or read? Usually the search topic relates to a certain city where the newspapers have been published or to a particular newspaper or a specific time period.

After the search, the user can select the desired search results and open those that seem to best match his/her search, based on the preview text closest to the search result. The search terms are highlighted to easily recall them. Users can also use fuzzy search, which makes the search more accessible and accurate, in the sense that word inflections and some optical character recognition (OCR) errors can be avoided and the user can find the information she looks for in a reliable fashion.

#### **3.2.2. Login**

The user has to login with one of the social media user credentials available. The login is needed because this is a way to collect the personal clipping book. In addition, it works as a simple protection against the not-so-friendly-users by inhibiting, e.g., the very basic types of spam comments.

#### **3.2.3. Clipping and Metadata**

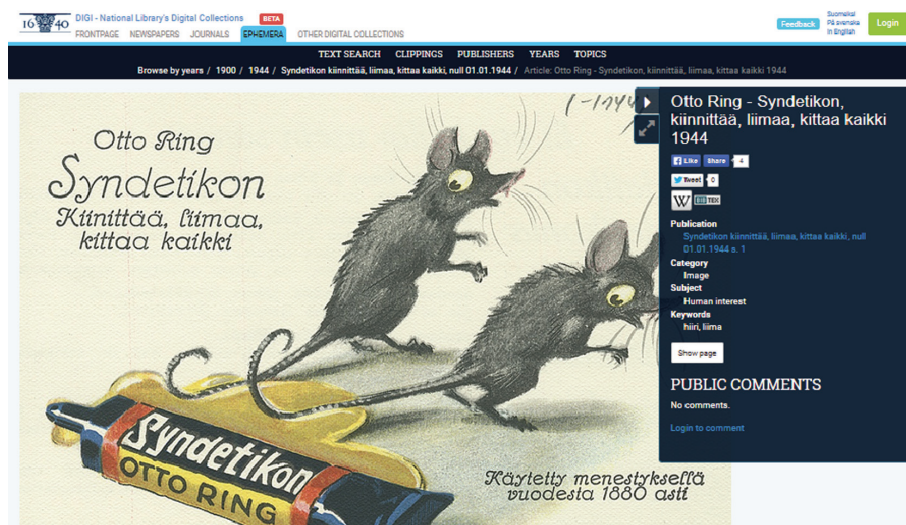
Clipping is done directly from the content page, for example a newspaper page. Users can drag the area they want, for example a continuous article spanning several pages via selecting the needed areas. Naturally the newspaper

metadata is already existing, but the end-user can add title, subject, topic and possible keywords to the clipping. The keywords are important as they make it possible for the user and for others to find relevant clippings. Based on our estimates, when testing the clippings workflow, on average it takes 2 to 10 minutes to create a clipping (depending on how long the clipping is, how many or how detailed the created keywords are, and how straightforward the selection of the pre-defined subjects and categories is). Naturally, after some initial learning period the activities become more effective. In addition, we have implemented a feature for the most advanced users which allows them to copy and paste the clipping metadata from one clipping to another. This helps those users, who systematically go through several bindings in finding their information.

### 3.2.4. Storing and Sharing

After the clipping is stored, the metadata becomes viewable and there is a unique link to the particular clipping. The end-user can share the clipping to social media, to Wikipedia or store the URL to email it to colleagues via the sharing functions as can be seen from Figure 2. The clipping is stored both in the personal scrapbook of the user and in the generic clipping collection where all the clippings go.

Fig. 2: An Example of a created clipping of a glue brand advertisement from <http://digi.kansalliskirjasto.fi>.





## 4. Crowdsourcing and Metrics

As in any crowdsourcing activity, there are both internal and external, tangible and intangible benefits, which are searched. The question of impact is also interesting — has enabling crowdsourcing shown an impact to some context, which can be specified and measured? All the metrics of this chapter are examined in regard of their purpose and of what kind of conclusions can be drawn from the first year of collected data.

### 4.1. Metric 1 — The Number of Clippings Over Time by User

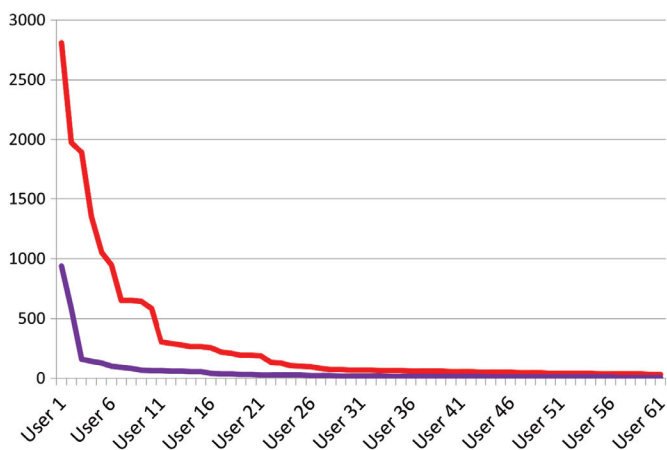
One of the most basic metrics is the number of clippings that has been made by a user. This shows his/her overall interest toward the crowdsourcing capabilities but it can also show how this interest changes over time. For example, we hypothesized that in the beginning the crowdsourcing metrics would start slow, but that they would increase over time, when a) there would be returning users, who had formed a habit of the clipping creation or b) when the general knowledge and communication about the usefulness and the fun of digging through the newspapers and creating clippings would become apparent.

The benefits of this metric is that our presentation system <http://digi.kansaliskirjasto.fi> stores the created clippings, with the date of the creation and all related data. So, with a simple database query the data can be collected and analysed.

In the beginning we also considered whether the 1% rule of the internet culture would be visible in the NLF crowdsourcing effort. The 1% rule states that in any internet community there are 1% of creators (highly active users who participate wholeheartedly), 9% of commentators, and 90% of spectators (who use the system, but not in a deeper way). This 1% rule can be interpreted in different ways: is it the percentage of all the page visitors who have registered, or is it the percentage of registered users who become very active? We used the latter approach in our metrics in order to follow how the new functionalities impact individual users.

After six and twelve months the results of the clippings quantity by user were as described in Figure 3.

Fig. 3: Clippings by user (6 and 12 months) (partial).



The purple line illustrates the situation after six months and the red line after twelve months. Points of interest are the most active users who have made nearly 1000 or nearly 3000 clippings respectively. This actually repeats the same finding as reported by Holley (2010a), that ‘super’ volunteers outshine the effort of others. So getting the few highly active users is very significant to the crowdsourcing as a whole in terms of quantity. The second point of interest is the midrange of users who have made a few hundred clippings. An interesting observation is that over the months this midrange has increased and expanded, which seems like a positive sign for the functionalities as a whole. Finally, the long tail of low-activity users is another main factor, and their amount has nearly tripled between the six month and the twelve month phase. There are a lot of users who might try the clipping once or twice: maybe they found the content they looked for and are happy with that, but they do not form a habit of using the crowdsourcing as such. The super users tend to follow a theme, when the single clipping makers might find one particular gem — but it requires a bit more research to see real patterns of usage.

#### 4.2. Metric 2 — Keyword Quantities

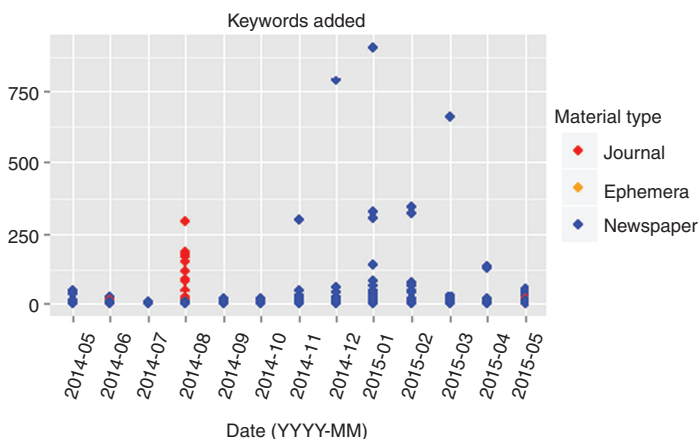
Another metric that we considered in order to understand the different usage purposes of the materials is the keyword or tag usage. This was studied by simply counting and adding the keyword quantities across the date when

the keywords were added, see Figure 4. In short, there are only a few very extensive clipping keyword clusters, but there is a long tail of more specific and rarer keywords also present here. The clusters of keywords tell about a specific theme that the users have created — it can either come from one or two active users, or it can be formed by several people who participate in the same theme.

In the keywords a blue dot represents a keyword in newspapers, red in journals and green in ephemera. As it can be seen from the graph, newspapers form a clear majority, superseding journals and ephemera. There is, however, also one interesting cluster for journals.

When checking the graph it is clear that there are some peaks for certain keywords. In December and January car related topics were searched, and they are highly visible in the keywords. A second main cluster is formed by the dance-theme: various dance events, dance schools that are visible in the digitized newspapers. The cluster of journals is mainly focused on word puzzles and quizzes, which were collected in August 2014. It is interesting to note that, like in the case of the tagging feature of the Historic Australian Newspapers (Holley, 2010b), also in our case the top users have made or added tags or keywords to nearly 1000 clippings for their theme of interest. As far as user amount goes, Australia and Finland are quite different due to the size of the language area (mainly Finnish and Swedish content compared

Fig. 4: Quantities of keywords over time.



to English), the profile of user interests seems to be in the same range even in this early phase of crowdsourcing.

There is also a constant usage of the keywords, so despite the fact that the users change and vary, they still follow the unwritten community rules and attempt to add at least one or a few keywords to the clippings they create. We have been happily surprised with the multitude of clippings and keywords made; people seem to be thoughtful with the keyword selection and it has highlighted the various user groups of the digitized collections. The clippings feature has given us insight into the users' interests: there are local historians focusing on certain regions, theme-based users who collect material, e.g. from dancing or from technological advancements, and, expectedly, family and history researchers.

One of the ideas was also to compare the user-picked keywords to the full listing of the keywords provided by the Finnish Ontology, which the <http://digi.nationallibrary.fi> service uses and to see whether users use those or develop their own ones. We have also noticed that the Finnish generic ontology has been utilized quite well by users and that its keywords are used. Users have found in this ontology the required concepts and this has maintained a good structure of the keywords in the clippings. In addition to ontology words, end-users can also define their own keywords and therefore new concepts were also created. The ontology gives structure to the tagging, but it might still be that in the long run some clean-up of typos, or singular/plural form corrections might be needed for the clippings. Nevertheless, even now any user can fix erroneous tags, which gives them the opportunity to participate in the creation of larger clipping collections by creating, updating or deleting keywords. From the administration side, however, it would be nice to have a tool to do simple tag fixes for multiple clippings.

#### **4.3. Metric 3 — user feedback analysis**

One thing on which we also kept an eye for was the user feedback that we received after launching the new version. At the beginning we decided to wait and see what kind of feedback we got, so that we could prepare the necessary user manuals, the frequently asked questions and the remaining appropriate help material.

The changed search facility was clearly one thing that has shown up in the feedback. There are those who really like the changes and those who were quite reserved towards the implemented changes. In order to help with the search a new help page and instructional videos were created and the feedback was answered by giving example custom searches that solved the particular questions of a user. In fact, after adding a link to the frequently asked questions (FAQ), the amount of search related questions has gone down, so probably the FAQ and the added help materials have been useful.

Another thing extracted from the feedback has been that some participants have contacted us and told us about the things they have looked for and why. This has partly been because they have a certain feature request in their mind for which they want to give us background information or because they would like to get access to more materials via the public web. Partially it has also been pure enthusiasm where they have wanted to share their findings or their work done with the library. The open website shows material until 1910 and when people doing research reach that year, they wonder where they should seek for additional material — one alternative being, e.g., the legal deposit libraries.

## 5. Conclusions and Future Work

What can be seen in a crowdsourcing effort like Australian Trove (Holley, 2010b), or even Wikipedia (Wikipedia Statistics, 2015), or in our case is that getting up to speed with crowdsourcing takes a bit of time. Despite using newsletters and social media channels of the National Library of Finland, getting new or existing users to start using the new features happens gradually. Listening to user feedback is important as that can help in developing features which users find usable — feedback is valuable when prioritizing new development efforts.

The metrics also change based on the feedback. Originally we were interested in seeing if the clippings themselves would start to have their own interest for the users and, based on the metrics, there is evidence that this has happened and people are committed to do their research by using the clippings. It is possible to identify specific theme collections by keyword amounts that focus on certain key areas. It is interesting to see whether these major collections become more popular when more people stumble upon them. In that case, it might also be possible to utilize those annotations in helping with machine learning research.

There is also a long tail of users, who might make one or two clippings and who in this way accomplish their goal. Seeing the daily influx of clippings gives NLF visibility to the end-users — not just what they say they do, but what really happens with the presentation system and the digitized content. Within the presentation system the overall situation can be monitored locally, because otherwise the content could end up in numerous social media citations, where the usage does not become visible if the end user does not use citing or does not mention the source. We also need to follow the user feedbacks and utilize them when developing future enhancements. It is definitely a good practice to ask for user feedback and see how changes implemented might impact the metrics.

The metrics have been one way to monitor the usage of the clippings and it seems that more metrics are needed. One idea might be to do month-by-month or year-by-year comparison to see if we can find any cyclic variance in the user activity for any specific time period or event. At the next phase, the second thing would be to look deeper into the content: which content is used the most? The metrics could be monitored both from the viewing side and from the creation of clippings, and this could help us create a recommendation engine for new users. This way we could steer them directly towards potentially appealing content, from where they could continue according to their own interests. In any case, monitoring and creating metrics requires a continuous effort. The metrics used should have a clear goal and be monitored, so that they can be acted upon.

## 6. Acknowledgements

- Big thanks to the hard-working colleagues in the Centre for Preservation and Digitisation of the National Library of Finland.
- This work has been partly funded by the EU Commission through its European Regional Development Fund, the program Leverage from the EU 2014–2020.



## References

- Bernstein, J. (2006). Born digital, not yesterday: next-generation web user seeks interactivity. *EContent-Digital Content Strategies and Resources*, 29(4), 14–15. Retrieved July 22, 2015, from <http://www.econtentmag.com/Articles/News/News-Feature/Born-Digital-Not-Yesterday-Next-generation-Web-User-Seeks-Interactivity-15480.htm>.
- Bremer-Laamanen, M. -L. (2014). In the spotlight for crowdsourcing. *Scandinavian Librarian Quarterly*, 47(1), 18–21. Retrieved July 22, 2015, from <http://slq.nu/?article=volume-47-no-1-2014-11>.
- Cole, R. E. (2002). From continuous improvement to continuous innovation. *Total Quality Management*, 13(8), 1051–1056. doi: 10.1080/0954412020000001. Retrieved July 22, 2015, from [http://asq.org/pub/qmj/past/vol8\\_issue4/cole.html](http://asq.org/pub/qmj/past/vol8_issue4/cole.html).
- DigiTalkoot. (n.d.). Retrieved July 22, 2015, from [http://www.digitalkoot.fi/index\\_en.html](http://www.digitalkoot.fi/index_en.html).
- Duncum, P. (2013). Youth's remix culture off and on line. *Australian Art Education*, 35(1/2), 10.
- Holley, R. (2010a). Crowdsourcing: How and why should libraries do it? *D-Lib Magazine*, 16(3/4). Retrieved July 22, 2015, from <http://doi.org/10.1045/march2010-holley>.
- Holley, R. (2010b). Tagging full text searchable articles: An overview of social tagging activity in historic Australian newspapers August 2008–August 2009. *D-Lib Magazine*, 16(1/2). Retrieved July 22, 2015, from <http://doi.org/10.1045/january2010-holley>.
- Kettunen, K., Honkela, T., Lindén, K., Kauppinen, P., Pääkkönen, T., & Kervinen, J. (2014). *Analyzing and improving the quality of a historical news collection using language technology and statistical machine learning methods*. Presented at the IFLA World Library and Information Congress (WLIC) 2014 Satellite Meeting, 13–14 August 2014, Geneva, Switzerland. Retrieved July 22, 2015, from [http://www.ifla.org/files/assets/newspapers/Geneva\\_2014/s6-honkela-en.pdf](http://www.ifla.org/files/assets/newspapers/Geneva_2014/s6-honkela-en.pdf).
- Sokovic, M., Pavletic, D. & Kern Pipan, K. (2010). Quality Improvement Methodologies – PDCA Cycle, RADAR Matrix, DMAIC and DFSS. World Academy of Materials and Manufacturing Engineering. Retrieved July 29, 2015, from [http://www.journalamme.org/papers\\_vol43\\_1/43155.pdf](http://www.journalamme.org/papers_vol43_1/43155.pdf).
- Wikipedia Statistics – Charts – Finnish. (2015). Retrieved July 22, 2015, from <http://stats.wikimedia.org/EN/ChartsWikipediaFI.htm>.