Increasing Availability, Data Privacy and Copyrights of Digital Content via a Pilot Project of the National Library of Finland

Tuula Pääkkönen
National Library of Finland
tuula.paakkonen@helsinki.fi

Abstract

Availability, data privacy and copyrights are topics which the National Library of Finland (NLF) is currently working on to increase the availability of digital content. Currently, in Finland, the digital collection of newspapers, journals and ephemera at the digi.kansalliskirjasto.fi service is available for everyone, where, e.g. newspapers are available for the years 1771–1910. The material after that period is visible in the six legal deposit libraries all over Finland. Researchers, especially, but also other customer segments are demanding better access to digital content. We at the NLF are searching for ways to open the collections to wider audiences. Therefore, we have established a pilot project, where we are experimenting with new modes of operation in opening digital content for education and research. In the pilot project, Aviisi, the participants are media houses, copyright organisations, universities, schools, public libraries, archives, museums and research organisations. We are creating tailored contract models to improve our responsiveness to copyright and data privacy issues. Beside the contract modes, we are also working with the technical solutions and communicating with the participants about possible ways to utilise digital content. Our objective is that, by the end of the project, we will have clarified legal, technical and availability questions, so that we will be prepared for most common scenarios and can proceed with the pilot projects, and then onwards after the overall success of all our pilot projects. Currently, we have fifteen pilot projects (hereafter, pilots) running, which cover about 40,000 people in the Mikkeli region, where about 4,000 of those people are...
Increasing Availability, Data Privacy and Copyrights of Digital Content via a Pilot Project

directly involved in the pilots as an estimate of students, teachers, researchers or other contact points. In this paper, we will discuss the pilots, how we have approached them to use digital content and attempted to increase the availability to the content overall while honouring the regulations of data privacy and the rights of copyright holders. In the last few months of 2016, as we finalise the project, we will aim to take the learning forward into working models, which will pave the way to further expanding the concept.

Key Words: digital collections; research use; availability; case study; newspapers

1. Introduction

One of the constant queries by the clients of the National Library of Finland (NLF) is how they can access specific copyrighted newspapers. There are two solutions, which the NLF currently has available. Firstly, we instruct the client to visit the six legal deposit libraries in different parts of Finland (National Library of Finland, 2016). The second alternative is to encourage the use of microfilm, which the client can request via the public library. Alternatively, gaining access to newer materials requires that the user travels to their nearest legal deposit library which can be several hundreds of kilometres away, depending on clients location.

In the earlier projects of the Centre for Digitisation and Conservation of the National Library of Finland (NLF) the focus has been on getting born-digital content delivered in digital form (Sorjonen, 2011), machine learning to improve digital materials (Digra-project) and utilising digital materials in new ways, such as the “Real Case Lab” project (Kapiainen-Heiskanen, 2012) and the Digibus-project (Digibus, 2015). There has been work undertaken with the copyrights in other parts of the NLF, but not as a full regional project in collaboration with the news media houses, copyright organisation and education and research organisations. All of the organisations, and their staff and clients, act as a testing site for Aviisi, where especially education and research are our points of focus.

The Aviisi-project aims to find ways to extend the availability for both education and research purposes via pilots (Karppinen, 2015). There are fourteen pilot participants (Figure 1), which are from different education and research
institutes. Pilots have the support from two news media houses and the Finnish Copyright Society (Kopiosto) in order to offer a full range of content to the users. In the thirteen pilot organisations of the Aviisi-project, the employees, students and customers of the organisations, the potential user base, consists of about 4,000 people.
The responsibilities and the rights of the pilot participants need to be stated clearly and for us the correct place was in the contracts. The contract structures enable new kinds of use, but they also contain the required guidelines about what users can do with the materials. The parallel copyright clauses, e.g. exceptions for the text and data mining, enable a wider variety of how digital content can be used in research.

We have also been analysing the impact of the updates of the European Union’s General Data Protection Regulation (GDPR) in the Finnish context. The Finnish analysis was undertaken as part of the creation of a first analytic report by the legal group of Aviisi, see Salokannel (2016b). The report investigates how the EU data act impacts the contractual space in research and education.

Besides the contracts, technical changes were required, such as better user access controls and usage reporting, which were already identified during the creation of the contracts. Naturally, as one of the permanent objectives, the usability has also been one of the focal points via more user-facing changes. In addition, for the end-user we have improved the search functionalities and created new ways of, for example, how researchers can obtain and analyse the digital content in a better manner.

1.1. Newspaper Digitization and Accessibility in Europe

In the European Commission (2016) report on digitisation, online accessibility and digital preservation is mentioned and the costs of digitisation are estimated to be 100 billion euro and would take around ten years. Therefore for that long-term cost, also the accessibility needs additional focus. At a European level there has been efforts to increase usage of digitised public domain material e.g. via dedicated portals, hackathons or tools (European Commission, 2016) which both increase awareness but also accessibility to the materials. For the in-copyright material there are specific solutions for the orphan works and online use of digitised contents. For the orphan works there in several member states, there is national legislation in place where the aim is via diligent search to find the rightsholders and after that there are specific provisions allowing also use of the orphan work, e.g. in Finland for the goals of the public good in certain institutions (European Commission, 2015; Act on the use of orphan works 764/2013).
For the print content, mainly newspapers and journals like in the case of the National Library of Finland, our approach has been via the contract model. Many Nordic countries have the Extended Collective Licensing (ECL) system in use, in order to support the mass use (Vuopala, 2013). This method has allowed for example the cross-border collaboration of the National Library of Sweden and Åbo Akademi University in Finland, so that the university researchers can have access to the materials of KB remotely (Kungliga Biblioteket, 2015). Our approach in the Aviisi project is similar, having collaboration between the local public libraries, museums and education institutes of all levels in one region.

2. Availability

The goal of the Aviisi project is to bring the golden century of newspapers to new usage purposes. The approach chosen was to increase the availability of the newspapers via multiple pilots. The pilots targeted different levels of education, research and archiving within institutions, with the hope of gaining an understanding of the possibilities of utilising digital material in these contexts.

2.1. Local Pilots of the Aviisi Project

The pilot population within Aviisi is quite extensive. In the key role were the publishers, Kaakon Viestintä and Viestilehdet, who graciously approved the use of their newspapers, Länsi-Savo for the period of 1916–2013 and Maaseudun Tulevaisuus for the period of 1917–2013, in order that they could act as the content for the pilots. The key here was that it was possible to use their whole publication history, which gives unique opportunities to research that time-span of history. As aforementioned, the whole archive of the newspaper is only available at the six legal deposit libraries, hence Aviisi had quite an impact in this case.

Also the user organisations of the pilots were very versatile. For a generic grouping, the pilots can be divided into a) teaching and education, b) research organisations, c) archives and d) museums, as illustrated in Figure 2. In fact, to get all schools to participate, we made an agreement with the city of Mikkeli, which in the end covered schools, museums and public libraries. Then there were agreements made with the Central Archives for Finnish...
Business Records and the Provincial Archives of Mikkeli. For the research-oriented pilots we were able to expand beyond Mikkeli, namely to the Folklife Archives of the University of Tampere. From the University of Helsinki, the whole of the National Library of Finland, Ruralia Institute and the Fin-Clarin consortium are also full pilot participants. With careful planning and negotiations, even before the Aviisi project was started, we were able to achieve this wide range of pilots.

The first objective in the Aviisi project was to gather information about potential pilots, setting up contacts, possible IP addresses for creating the connections and having a system of communications ready. We started, firstly, with the leadership of each organisation, to bring them up-to-date with the expectations of the pilot and also because in the Aviisi project we would know what each participant desired from the pilots. With respect to the leadership, we approached the core personnel, who in most cases were given a short introduction to the materials or at least to the material on the introductory page they received. This approach raised management approval, and gave enough information for the local experts, so that they were ready to be consulted by their clients or by others.

2.2. Scope of EU General Data Protection Regulation and Copyright Directive

Currently in the van of development in EU regulations that are of interest to the digitized collections are the EU General Data Protection Regulation

Fig. 2: Pilots of the Aviisi project.
(GDPR) and the incoming copyright directive (also known as DSM-directive). The former focuses on the protection of personal data of natural persons. In addition, the GDPR enables persons to check their information and rectify their information (Article 4, 16) and data in general can be utilized for purposes laid down by law (Article 8, 12). With regard to these, the data subject should have contact information for rectifications as well as about how long and where the material is stored, for example (Article 13, 14, 30). There are also organizational requirements via monitoring and communicating any possible personal data breaches and assigning a person for monitoring organization’s compliancy towards the data protection directive (Article 28, 32, 33, 34), logging of certain processing operations occurring to the materials (Article 25) and security of processing of personal data (Article 29). All of these also relate to varying degrees to the established good principles of a development of an information system.

The proposed copyright directive, which is also known as Digital Single Market (DSM-directive) is then the second directive, with new requirements towards digital collections. Firstly the Article 3 allows text and data mining in scientific research to works where there is lawful access, this is very relevant to libraries as libraries can take a more prominent role as data providers for researchers. Secondly, Article 5, permits cultural heritage institutions to make copies for preservation purposes – for example, this kind of legislation has already been in place, which has been useful for example in digitising most-used materials in order to preserve the originals. The third point in the DSM-directive is the ensuring of the compensation of the authors either via publisher or so called information society service providers (Articles 12, 13).

All Galleries, Libraries, Archives and Museums (GLAM-sector) are impacted by these regulations as active cultural heritage organizations. For example, in Finland there are work groups for considering which parts might require local adaptations in order to protect this relatively small language area where the markets are quite small. The Ministry of Education has also lately requested comments from various organizations in order to get a broad view about the copyright legislation – this information can be used in communication towards EU and also when considering how and what is included in the local legislation.
The challenge why we have to work both with the GDPR and the copyright legislation with regard to the digitised newspapers is related to the content and the publication time of the materials. Newspapers can contain personal information, either given freely in an interview or published in a family announcement. Then on the other hand, material is not copyright free until 70 years after the authors’ death. For the in-copyright material, Kopiosto (the Copyright Society for authors, publishers and performing artists) is the negotiation party of most of the copyright agreements of newspapers (Ministry of Education, n.d.), and the latest newspapers are negotiated directly with the media houses. Because in Aviisi the pilot’s target was to get to as recent material as possible, it required contracts with both Kopiosto, media houses, the library itself and pilot users. Already in the first discussions both the data privacy and copyright issues came to discussions as the usage purpose has significance in possibilities of extending the access to the materials, so these two aspects become intertwined. So, when the National Library of Finland wants to offer more materials to the users, we need to prepare for the copyrights and data privacy, because the general public and the authors should be able to trust us to comply with the regulations, which should be balanced with the aim to get more material to be accessible for more usage purposes, which is indicated by the multitude of pilots. A new regulation and directive comes in force within two years after it has been published in the Official Journal of the European Union, so with the Aviisi pilot users, it was a suitable time to start thinking of impacts early, to develop and test new contracts, processes and system features. This gives us a head start and gives us the opportunity to take a long-term view to the future development requirements.

3. Data Privacy

The incoming EU General Data Protection Regulation (GDPR) was one element that we wanted to look into more deeply in the context of Aviisi. Therefore, after some discussion in the legal group, it was decided to produce a separate report by commissioning the work to an external expert. The external expert enabled our having a much broader view of the data privacy issues, while keeping the report contents independent from the views of those involved in the Aviisi project.
3.1. Finnish Analysis of EU Data Privacy Law Analysis

The timing of the data privacy report coincided with the update of the EUGDPR, i.e. The European Union’s General Data Protection Regulation 2016/679, which came into force in May of 2016 and will be applied from 25th May 2018, (EUR-Lex, 2016; Hunton & Williams, 2016). The report on data privacy had specific exceptions for the newspaper archives, which were seen as positive indicators for the context of the Aviisi pilots (Salokannel, 2016a). As the NLF has a legal obligation to store materials and to make them available for use by the general public, this also, by the nature of the materials, makes it possible to process information which might relate to persons within the digital context. Also, because of the nature of the material, there has to be a search function in the materials (Salokannel, 2016b), otherwise, services to the users would suffer. However, for this kind of use, there must be both a legal and a technical setup in place, which respects both data privacy and copyright requirements. It is yet open to debate whether historical newspaper archives could also utilise the journalistic exception in the context of own use by the journalists of the media house (Salokannel, 2016b), as it might not be applicable in this context (Heikkinen, 2016). One of the key open questions is related to what the tasks of the library are meant to encompass. In that, currently, can ‘availability’ be interpreted to mean digital, which is the main source of use (National Library of Finland, 2015), or does it still mean actual physical visits to the traditional library? The digital form enables availability in unforeseen ways, regardless of time and place, and to all different users in an equal manner.

The second most discussed topic with regard to the EUGDPR, is the “right to be forgotten” ruling. It currently appears to be in a two-fold situation, and having a national legislation and discussion in place could clarify the situation. In one court case in Belgium in 2016, it was decided that even the archives of the newspapers themselves might be such that the right to be forgotten ruling applies (Belga, 2016). However, in this case the reasoning was mainly due to the current Belgian data privacy law whereby an individual would suffer damages (such as in a car accident in the 1990s), which was seen as being of higher importance than freedom of speech and of the press. However, in another case, the French High Court decided in the case of a particular newspaper archive, that freedom of the press is a more powerful right, than the right to be forgotten (Van Quathem, 2016), when the case was about
sanction information from the year 2006. Hence, the cases are different, and it might even be that only over time a suitable balance is found. The different cases might require case-by-case analyses (Heikkinen, 2016), whereby every situation is analysed based on the factors impacting it.

3.2. Technical Impacts

The EU GDPR, has quite wide exceptions for the archiving purposes in the public interest (for example, sections 50 and 65), and similarly for scientific and historical research (EU 2016/679). The EU regulation is binding and applied as-is, but there is also a directive, whereby member states can define their own laws to achieve the goals of the directive (European Union, 2010). In a way, as creation of information systems takes development over time, the work should be started quite early on, in order to be ready when the regulation and directive come into effect.

For example, in the Aviisi project, we have started to work on the possibility of redacting certain material from the collections based on the specific, detailed request by the rightsholder. The redaction of the content (for example one article) is only ‘in effect’ only there where the availability to the digitized materials has been extended via contracts. This means that in those places where there are is a legal obligation to make material available, like the dedicated terminals at the legal deposit libraries, the original material is shown as usual. From the technical point of view, different rules for legal deposit libraries mean some amount of extra effort and consideration in defining the process and the requirements. The hiding requirement comes mostly from the copyright needs, because then it is a more granular way to remove material if a request arrives to the copyright organisation and is then forwarded to the National Library.

The second main object of focus, in a technical sense, would be the access control mechanisms. It seems that the trend is to go to even more exact access limitations, for example from network address-based access to the individual accounts. Technically, for example, a national identity federation service could be one way to give access to certain materials via the identity of the university or research institution. For any more granular access control, there would need to be an approval processes, whereby the authorisation is given
to those who are especially allowed access. With approval, the home organisation can take care of the user support of that home organisation and from the point of view of the library, the responsibility of the user is with the home organisation.

4. Copyrights

Currently, the material which we can show in the digital collections of the National Library of Finland is limited to the year 1910. Therefore, as in the Aviisi pilots we have been opening up materials to the pilot participants up to year 2013, we have made a special effort to educate both pilot users and organisations, so that they realise that they have access to material that is rarely seen.

To clarify possible copyright issues, we have had a separate legal work group within Aviisi. There have been participants from the Finnish copyrights organisation (for example, newspapers), Kopiosto. In the view of the Kopiosto the agreements are seen as a way forward, when finding ways to agree with all stakeholders (Kingsley, 2015). In the group, there is also a lawyer from the National Library of Finland and a lawyer from the University who, together with the Aviisi project manager, creates a compact group, to discuss copyright topics within the context of these pilots. Also, when the legal group discussed the data privacy topics, there was an academic expert of information law present in some of the legal group meetings, who also gave valuable insights in the discussions.

4.1. Copyrights and Contracts

One solution to the copyrights is actually via contracts. In each of the contracts made, we have specially mentioned what can be done within the pilot organisation, and also what the expectations for the users are. In short, the pilot organisation is allowed to use the pilot materials for research and education. It is not allowed to share material beyond the pilot location, for example, with social media. Based on the Finnish copyright law 12 §, it is however possible to create copies of a copyrighted work for private use. For example, in the community college course this was a pathway that enabled
course participants to get copies from the class usage that would be used at a later date.

Widmark, Holm and Reilly (2016) claim that “current copyright laws are no longer optimal” and their examples circle around the needs of having information available anywhere (without dedicated terminals or the limits of national borders). In Finland, legislation allows undertaking contracts and agreeing on the rules, together with all relevant parties. As libraries do not have any commercial interest and they have their own specific goals (Widmark et al., 2016), this gives libraries an opportunity to take the role of an impartial centre in connecting users and copyright holders. With well-considered contract models, the whole negotiation process can be made efficient, even if there would be hundreds of communities or schools and multiple publishers. However, the whole contract management lifecycle from the start of the negotiation to the completion, requires time and effort, and the more variance there is in the offered content packages, the more business acumen and legal expertise is needed. As with any tool, creation of contracts is not without cost, but it might still be more flexible than possible changes in copyright law.

Together with the contract model, also the cost structure still requires work on both sides. The work needed to clarify the copyright owners of all of the articles of all newspapers, throughout all years, is extensive. There is also a need to combine that information with proprietary information of copyright organisations, which means that finding ways to automate would necessitate new kinds of collaboration. One issue with possible cost structure models is scaling – the amount of pages compared to the actual users has to be feasible in order to be negotiated with the funders. Naturally, besides contracts, during Aviisi we created additional ways to both manage the access to the contents and also to monitor the usage in more detail. This was both to enable better statistics and auditing that the opened material is used only from the agreed pilot locations.

4.2. Copyrights and Technical Controls

In the legal group, there was actually an abundance of discussions about risks. No matter what kind of technical solutions are done to the presentation systems of materials, there are always new tools and tricks appearing, which
can find ways to circumvent any previously made restrictions. For example, in the legal deposit libraries it is possible to print a page for oneself, but the question is what happens to that paper document afterwards. If people are organising their lives with respect to the Online Cloud services or personal archives, is it then expected that they scan the material again and upload it to their own Cloud service? Is that then a problem?

Another topic that was discussed in the Aviisi legal group concerned the digital copies and the sharing of them. Figure 3 shows a categorisation of computer set-ups of pilots, that we have seen during the Aviisi project. If we begin from the legal deposit library computers, they are the most constrained environment. There, the usage is limited to specific dates and times (namely, when the library is open), and it is possible to view and print materials or to photograph them from the screen (screen shots or digital camera). The researcher or customer computers are those which a certain pilot organisation offers for its clients, for example in a study room, where there might also be access to the other content services. Then, the organisation-owned computers are like in the case of the Ruralia Research Institute, where the computer is owned by the University of Helsinki and where the usage policy defines its

Fig. 3: Computer capabilities within the pilots.
usage for the work purposes. The 4th layer computers are the personal computers of end-users, where they can be adhering to any possible setup. That case would have been resolved when, for example, the public library authentication system would have been implemented.

4.3. Copyrights and Metrics

Besides the annual reporting requirements, the contract-based approach of increasing availability creates new requirements for the reporting. The copyright holders and publishers want to know how much their content is used, either for compensation or for the prevention of unauthorised use (Rautiainen, 2016). In fact, with the discussion of these parties, both are generally curious about the usage of content, for example, the annual reports do not go to the level of the individual publisher.

The second group that would like to see the metrics information naturally consists of the pilot organisations themselves. We have fourteen pilots running, and they also would like to know how their usage scales compared with the totals. There also seems to be a case of when there is an active key user in a pilot organisation, they also want to know how well their own internal promoting of the Aviisi project is doing. The current hypothesis is, that the smaller pilot organisations are directly investing their efforts with respect to the benefits they see directly. However, the bigger organisations, where there are many end-users, would require more communications tools and support from the Aviisi project. In fact, in the beginning there were ideas about having joint workshops to design communication methods, but unfortunately the schedules were never met. During the period of late May until early June 2016, the Aviisi project undertook a survey about the usage of the materials in schools, where one key point was also to evaluate the views within Aviisi with respect to the other users of Digi.

The third group, which needs statistics and metrics, is the Aviisi project and the National Library of Finland itself. The annual report covers metrics, page views and content, but it could be seen as beneficial to go deeper into the metrics. For example, in a recent study, where the National Archives and the National Library of Finland usage patterns of users were analysed, it was recommended to segment the users in a better manner in order to understand
the content and service needs (Hölttä, 2016). This is in line with the views of Tanner (2013), whose balanced value impact model could be something to be experimented with, when we start to undertake a more deeper analysis of the impact which the Aviisi project has had in the Mikkeli region, in the NLF itself and in publishers and pilot participants.

5. Conclusions

The availability, data privacy and copyrights of digital data were focal points in the Aviisi working groups. The working groups planned the pilots and considered the possible problems, which then were taken into account either in contacts, technical solutions or via communications with pilot users.

As noted by Tanner (2013) and Hölttä (2016), libraries need to discover and to define who are the users of their digital materials. Contact with real users was truly beneficial within the Aviisi project. The direct contact with users has shown us at the library how the digital materials and information systems were used and what are the requirements of the users. The user survey about how the materials were used was undertaken in the early summer of 2016 and will be analysed in the final phase of the project.

Already now it is still good to note that piloting these new kind of usage models has been beneficial and could be a useful model for other libraries to use. Via piloting all different stakeholders can experiment with the implications of enabling new access methods to the materials, users get a glimpse of what is possible and the publishers can see the potential interest for the content. In addition, via pilots, there is an opportunity for a dialogue to discuss and find common solutions to the questions of open access or at least extended access to the digitized materials. It should be possible to find common ground, where every participant benefits.

In the future, we will continue with collaborations via building up contacts with news media houses, universities and schools which lie beyond our current pilot area. The current pilot area, in the Mikkeli region, has been a great area for experimentation, and we believe that we achieved realistic results from the pilots with regard to availability, legal and copyright issues. The
Finnish implementation of the copyright and data privacy laws requires preparation, but we do believe that the pilots and the Aviisi project as a whole have helped us at the Centre for Digitisation and Preservation to prepare for these changes.

Acknowledgements

Aviisi project and this work was funded by the EU commission through its European Regional Development Fund and the program Leverage from the EU 2014–2020.

The author would also like to thank all of the colleagues in the National Library of Finland and the reviewers of LQ for their good comments.

References


Web sites referred to in the text