

## CHAPTER 10

# Of Great Men and Eurovision Songs

## Studying the Finnish Audio-Visual Heritage through NER-based Analysis on Metadata<sup>1</sup>

Maiju Kannisto and Pekka Kauppinen

### Introduction

Part of a nation's cultural heritage is produced and preserved through audio-visual archives. Whatever has made it into the archival collection has passed severe processes of selection, which secure for certain cultural products a place in the cultural memory of a society. This process is called canonisation.<sup>2</sup> In this chapter, we explore how the canon is built up in the national audio-visual archive in the digital age. We study this by asking which people and what events and periods are 'remembered' not by the members of a nation, but by a collective national memory resource, the online audio-visual archive of a national public service broadcaster. Like particular natural and historic sites considered as 'heritage', and thus referring to a cultural and historical resource for all generations, past radio and television is now protected through copyright and continued cultural recirculation and exploited as both private and public property.<sup>3</sup> Our study focuses on the Finnish public service broadcasting company

---

#### How to cite this book chapter:

Kannisto, M., & Kauppinen, P. (2020). Of great men and Eurovision songs: Studying the Finnish audio-visual heritage through NER-based analysis on metadata. In M. Fridlund, M. Oiva, & P. Paju (Eds.), *Digital histories: Emergent approaches within the new digital history* (pp. 165–180). Helsinki: Helsinki University Press. <https://doi.org/10.33134/HUP-5-10>

Yle (former Yleisradio, founded in 1926) and its online archive. The dataset used in our research consists of Yle's archival metadata, which we analyse as a historical source material using the method of Named Entity Recognition (NER) as it is implemented in the digital tool the Finnish rule-based named-entity recogniser (FiNER).

Yle's online archive, The Living Archive (*Elävä arkisto*), presents part of Finland's audio-visual history. The archive is an illuminating case study as it is a large editorial historical audio-visual service with open access metadata, as well as a high profile among Yle's media output. It was first launched in 2006 on the 80th birthday of the Yle Company and the Living Archive enabled Yle to celebrate its history, while at the same time providing the public with new ways of watching television programmes online. The core idea behind the Living Archive is that audio-visual clips are historical source materials—documents from and about the national past representing Finland's national audio-visual heritage. In this way, the archive contributes the nation-building process by formulating collective national identity, as suggested by Benedict Anderson in his theory of nations as imagined communities.<sup>4</sup> As media researcher Derek Kompare suggests, the audio-visual heritage serves as a base of legitimacy for audio-visual media and memory; namely, as something worthy of attention, preservation and tribute. It can be used as a cultural touchstone, instantly signifying particular times.<sup>5</sup>

The Living Archive continually publishes new material from Yle's archives for viewing and listening via the online service. What is published is based on topicality, new copyright licences, Yle's current programming strategies, audience wishes and new archival discoveries.<sup>6</sup> Archival material has accumulated from the first audio recordings at the beginning of the 20th century to the present: Yle selects and presents the audio-visual archive material for current users in constantly new ways by adding and framing the material in 'background articles', written by journalists and archivists. Media history researcher Mari Pajala has analysed the ways in which the Living Archive attempts to make the material 'alive' and meaningful in the present through its journalistic front page, background articles and the possibility for interaction. In this way, by tying in moments of archive television with current events and television programmes, the archive continually connects the present with the past.<sup>7</sup>

The aim of this chapter is to produce new knowledge on the canon built in the Living Archive. However, at the same time, metadata as a source material and name recognition as a tool in a historical study must be analysed because the canon is connected with the limits and possibilities offered by the digital material and tool. In this chapter, we first introduce the research material and our method, the metadata used and the NER-based analysis. This analysis, using the FiNER tool, enables the identification of particular historical individuals, events and years from the metadata material. Finally, we discuss the limits and possibilities of our research process and results.

## Finding Voices and Images of the Finnish Past: Metadata and FiNER Analysis

So far, since 2006, tens of thousands of audio-visual clips have been published in the Living Archive, so a computer-assisted method is needed to make larger sense of such a vast source material. The historical researcher's role in delimiting and selecting relevant material for analysis is crucial to the success of the computer-assisted analysis results. In this chapter, the research data does not consist of the media clips of the media files stored in the Living Archive as such, but of the metadata of the media information of these media files. Yle made this metadata material available on 3 January 2018.<sup>8</sup> Of this metadata, we selected the columns describing the title of the media, the promotion title and the description of the content (in Finnish). These columns describe the content of media clips, but not in general the author information. We did not want to include the latter in this study as we preferred to focus on the constructed national heritage and the individuals presented therein—not the authors of the various media documents. This could, however, be a relevant topic for future historical research.

The use of metadata as data is problematic in some respects because the material is inconsistent and contains some double data and gaps, as well as, at times, false information. An essential part of the historical source criticism is thus to find out the classification and guidelines that have been used to create the metadata. In the Living Archive metadata, the media descriptions have been produced in three main ways. (1) Descriptions have been made by the authors of the original programmes in connection with their original production to serve as programme information for other media. (2) The reporter or administrator of the Living Archive has written a short description of the clip's content in its subject field. Since this description field has not been displayed to the end user, this field is not always completed, and thus about one-tenth of the fields are empty. (3) Due to technical failure in the migration of the archival clips in 2011 (the transfer from one technical platform to another), the media information on the clips' still images were incorrectly entered in the media clips' subject field.<sup>9</sup> Therefore, in place of the description of the audio-visual content, this field contains the content of the still image, together with the photographer's name. Only the content of some video clips published before 2011 may have been edited and supplemented after that year. In the analysis, we have been aware of the above-mentioned issues and taken them into account as much as possible (for example, we excluded the photographers from the analysis).

NER is a task in Information Extraction consisting in identifying and classifying some types of information elements, called Named Entities (NE). It is stated that NER analysis usually responds to the five typical questions in the journalism domain: what, where, when, who and why.<sup>10</sup> In our analysis, we utilised FiNER,<sup>11</sup> a rule-based NER tool loosely based on the Swedish NER

HFST-SweNER.<sup>12</sup> At the time of our study, FiNER had not yet been used extensively as a research tool and therefore this chapter is also partly a methodological experimental study.<sup>13</sup> FiNER was created for the FIN-CLARIN consortium at the University of Helsinki, which is the Finnish part of the European CLARIN (Common Language Resources and Technology Infrastructure) collaboration for developing research infrastructure for language-related resources in humanities and social sciences.

FiNER utilises the Helsinki Finite-State Toolkit and its implementation of the *pmatch* (partial string matching) function,<sup>14</sup> which allows the compilation and implementation of pattern-matching rules as computationally efficient finite-state transducers (FSTs). FiNER's pattern-matching rules employ a number of strategies in finding and disambiguation names, including hints in string structure (such as uppercase letters, affixation, etc.), collocations, runtime adaptation and gazetteers (lists of names). The FiNER tool accepts any plain-text input, but works best on running text that adheres to Standard Finnish spelling and typographic conventions. We used a set of UNIX text-processing utilities to extract relevant segments from the tabular metadata for tagging, as well as to filter out any superfluous data that might slow down or interfere with the NER process.

In order to calculate total frequencies for names in the data, each matched word segment in the output first had to be lemmatised (that is, reverted to its uninflected form). Thus, we had FiNER output lemma forms and morphological analyses for each word and created a Python script that extracted matched sequences and used morphological information to lemmatise each of them. Once all names had been printed out in their lemmatised forms, their frequencies in the data could be calculated with relative ease.

As digital history researchers Kimmo Elo and Olli Kleemola have pointed out, it is essential for the researcher to understand how the computer-assisted analysis produces the results.<sup>15</sup> Thus, we looked at the frequency lists of different categories of FiNER analysis from the perspective of the possibilities and limits set by the technical characteristics of the tool, as well as possible errors. After having removed the names that occurred as a result of technical error, we put together TOP10 and TOP20 lists from the names that have received most mentions in different categories. Finally, in order to understand the results, it was important to check the background articles in the Living Archive and to contextualise them tentatively with Finnish cultural and political perspectives.

### **Great Men, Journalists and Musicians as Remembered Persons**

People's names are generally better recognised than other name entity types.<sup>16</sup> FiNER recognised nearly 12,000 people in our research material; in the analysis, we focused on those who had received most mentions. When analysing the people, we found three main groups dominating the historical personage of the Living Archive: great men, journalists and musicians (see Table 10.1). Among

**Table 10.1:** Top 20 of the most frequently mentioned people in the metadata of the Living Archive listed by FiNER.

Number of mentions	Top 20 people	Their professions
353	Urho Kekkonen (includes 'Urho Kekkonen' and 'Kekkonen')	President of Finland
223	Arto Nyberg	journalist
125	Kirka	musician
111	Marjukka Havumäki	journalist
105	Vesa-Matti Loiri	musician, actor
90	Mauno Koivisto	President of Finland
89	Mikko Alatalo	musician, journalist, politician
89	Matti	
86	Bettina Sâgbom	journalist
78	Matti Rönkä	journalist
78	Danny	musician
76	Päivi Istala	journalist
74	Göran Palm	journalist
73	Mannerheim	President of Finland
71	Paula Koivuniemi	musician
71	Jaakko Selin	journalist
69	Lasse Mårtenson	composer, musician
69	Antti	
68	Susanna Ström-Wilkinson	journalist
64	Juha Laaksonen	journalist

Note: Photographers are excluded. Source: Author.

the Top 20 of the most represented people, President Urho Kekkonen is the most frequently mentioned person. Kekkonen (1900–1986) was the longest-serving President of Finland (1956–1982) and a politician who achieved an almost unchallengeable political position during his long presidency. President Kekkonen's prevalent role in the archival clips is also emphasised by the fact that his trusted photographer Kalle Kultala received the second most mentions in the material, although this is due to an incorrect overwrite in the media information. Kekkonen is the protagonist in many articles, which include a number of video and audio clips from election campaigns, presidential visits and public speeches, as well as clips of him carrying out his hobbies. During his 25 years of presidency, Kekkonen gave 25 presidential New Year's speeches and hosted 22 Independence Day receptions, both of which are essentially national

audio-visual broadcasts, repeated annually and among the most watched programmes on Finnish television.

President Kekkonen was connected to the Finnish Broadcasting Company Yleisradio in many ways. First, he had close personal relations with people working there. Second, his political career was concurrent with the development of broadcasting in the 1930s and 1940s. Finally, he often appealed to the people in his radio speeches during the growth in radio licences. It can be argued that his long reign was comparable to the monopoly of Yleisradio.<sup>17</sup> Media researchers Lotta Lounasmeri and Johanna Sumiala have argued that President Kekkonen used the new mass media skillfully. He managed to combine different roles by acting as a sovereign political leader in the mass media while, on the other hand, at other less formal occasions, performing as a man of the people by skiing, fishing and meeting people in the countryside.<sup>18</sup> Iconic images of the photogenic president were spread across the mass media, making these audio-visual presidential representations part of the nation's public memory. In this way, the Living Archive played an important but previously mainly neglected role in preserving and presenting a political part of the national audio-visual memory. Furthermore, President Kekkonen not only appears in the archival clips, but also as a character in documentaries, drama series and sketch shows, and as a reference point to which other politicians were reacting. This is not just a past but also a very active part of the national memory, as the cult of President Kekkonen still strongly survives as a nostalgia and longing for strong leadership.<sup>19</sup>

Among the Top 20 of the most mentioned people, there was also another President of Finland, Carl Gustaf Emil Mannerheim (1867–1951), who could be characterised as a great man in national history. Mannerheim, Marshal of Finland, was frequently mentioned, even though he appears in few contemporary audio-visual sources. However, he is retrospectively presented in the archive through later historical photographs, film clips, and radio speeches and documents. Furthermore, Mannerheim's character became a part of Finnish popular culture in the form of a controversial doll-animated short film in 2008 and through other films. According to historian Tuomas Tepora, the contradicting views have been an essential part of Mannerheim's mythical role in Finnish cultural memory.<sup>20</sup> His legacy also appears in the material when talking of The Mannerheim League for Child Welfare and the Knights of the Mannerheim Cross, awarded to Finnish soldiers.

Along with the great men, men in general dominate the media clips: 72 of the first 100 people are men. Of the 10 women mentioned, the top eight are Yle journalists; only singer Paula Koivuniemi and President Tarja Halonen are also cited. As is well known, in the Western historiographic tradition, women have rarely been treated as active actors until recent years and therefore they have not been given much space in national historiography. Although the tradition has been challenged for almost 50 years now, women are still under-represented in most canonised historical narratives.<sup>21</sup> This is also the case in the

Living Archive, in spite of equality work being in operation for a long time now. This is also related to the history of the feminisation of Finnish professions. President Tarja Halonen was the first female Finnish President (2000–2012). This can also be observed among the journalism profession, another dominant group in the archive. At Yleisradio in 1972, only 35% of the journalists were women,<sup>22</sup> and the number of female journalists only increased significantly from the 1970s onwards.

In general, the journalists of Yle to a great extent appear in the material. The Top 20 listing includes eight Yle journalists (see Table 10.1). In the metadata, journalists are mentioned in many respects. News anchors are mentioned in connection with the archival news clips, while journalists working on the Living Archive are mentioned in connection with the background articles. Journalist Arto Nyberg is the second most frequently cited person in the archive. He has since 2004 hosted a popular talk show named after him, where many Finnish public figures and celebrities have visited and many clips therefrom are published in the Living Archive. In this audio-visual history, journalists are the key figures and agents behind the production of the media clips and thus appear frequently in the material more in this capacity as facilitators and producers, rather than as objects of newsworthy events. This important factor of the media production system needs to be taken into account.

In addition to journalists and presidents, many musicians are also mentioned in the audio-visual archive: six are included in the Top 20 listing. What unites these musicians is great popularity, which often comprises long careers stretching to several decades of hit songs. When compared to the most frequently mentioned bands listed by FiNER, the musicians are more focused on older pop singers and evergreen hits, while the bands category is more versatile and focuses on newer bands. The most frequently mentioned artists are those who have had visibility in Yle's music programmes, such as Eurovision Song Contests and festival recordings. For example, Mikko Alatalo's prominence in the archive emanates from his versatile TV work in Yle's music programmes in the 1970s and 1980s.

### **Eurovision Song Contests, Sports and Wars Unite the Nation**

Music also plays an important role in FiNER's list of the most frequent events. In part, this is due to Yle's major role in recording music festivals and preserving them for the audio-visual heritage in the Living Archive. However, the largest national audio-visual event is not a national but a European event in the form of the Eurovision Song Contest. As an annual and long-lasting television event, the song contest has been an important event of popular culture. Historian Mari Pajala has argued that the Eurovision Song Contests have become a prominent part of Finnish national memory and history. The regular annual contests have particularly participated in Finnish nation-making by

creating media discussions surrounding the meaning of nationality in the collective cultural memory. The Living Archive reviews both the successes and the failures of Eurovision. Finland has historically been very unsuccessful in the contests, but this finally changed in 2006 with the hard rock band Lordi's victory. The long tradition of negative experiences and the decades of disappointment were then forgotten as public festivals reupted in market squares all over Finland.<sup>23</sup> In addition to the international contest, the national Eurovision qualifiers are presented extensively in the archive, which gives many artists a place in the national audio-visual history.

Another popular category of events is major sports competitions, such as the European and World Championships in ice hockey and track and field. The broadcasting of these major sport events has been strategically significant to the Finnish Broadcasting Company since they attract large audiences and are therefore legitimising the company's role as a public service broadcaster.

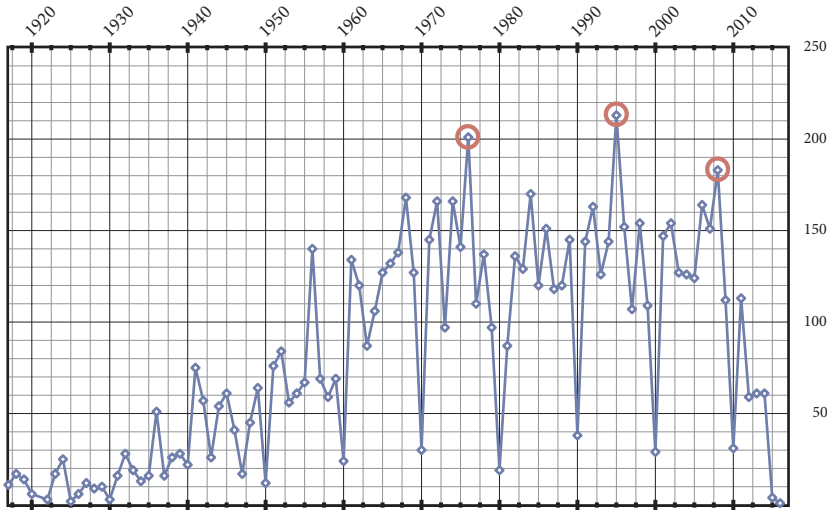
The third category relates to war. Wars and political conflicts identified by FiNER are listed as events, together with music, sport, art and entertainment, which can, however, be analytically difficult for comparisons. Among the wars, the Finnish Winter and Continuation Wars are particularly well represented in the media material. These wars are significant episodes in Finnish history and form part of the national mythology formulated through national publicity, as well as, for example, through their appearance in schoolbooks.<sup>24</sup> Therefore, these two wars also form an important part of Finnish cultural memory.<sup>25</sup> The Living Archive preserves the remembrance of wars in connection with different anniversaries.

According to Pajala, many of the traditional moments of the television year are explicitly related to nationality: Eurovision Song Contests and sport events unite the national public in excitement at the presence of their own representatives, while the state's sovereignty is celebrated at the President's Independence Day Reception, and in the national war films.<sup>26</sup> These national audio-visual events are accumulated and added to the Living Archive annually.

### Nationally Significant Periods as Recognised by Yle

Different periods from the 1960s to the 2010s are remembered relatively evenly in the Living Archive, even though most of the programmes have been preserved only in the last two decades (see Figure 10.1). However, it is important to note that in FiNER's analysis, the recognition of the decade numbers is a challenge: FiNER may not recognise these without using clues from the text contexts, and may instead interpret them as ordinary words rather than as markers for historical periods. This produces gaps in the time series. There are three peak years in the FiNER frequency list: 1976, 1995 and 2008. When tracing these years from the media material, we noticed that during these periods there were an extraordinary number of significant political and cultural events.





**Figure 10.1:** The occurrence frequencies of the years 1917–2017 in the time series identified by FiNER from Yle's metadata material. Note: The peak years 1976, 1995 and 2008 have been circled. Since the summer of 2015, all the programmes of the Living Archive have been published via the Areena platform, which causes a harsh decrease in the metadata material.<sup>28</sup> Source: Author.

We suggest that these events have become key experiences determining the past and are thus significant parts of the national memory.<sup>27</sup> These historical events are presented in the background articles and collectively remembered in the Living Archive.

Among the 1976 media clips, the long-distance runner Lasse Virén's wins in the 5,000- and 10,000-metre runs at the Montreal Olympics and the drama series *Myrskyluodon Maija* (*Maija of Myrskyluoto*), shown on Yle's main channel, are both highlighted as key collective experiences. The latter television series was set in the Finnish archipelago during the 19th century and pictured the life of the protagonist Maija and her family in six episodes. Lasse Mårtenson, who was among the Top 10 in FiNER's list of the most represented people, composed the music for the series. The series has been characterised as unforgettable by many, in large part due to the music. In actual fact, the piano transcription of the theme music is the best-selling Finnish music publication of all time.<sup>29</sup> The Living Archive contains several background articles about both the music and new versions of the songs, such as the highlights of 15 versions of the theme music performed by various artists.<sup>30</sup> As evidenced by its many occurrences in our source material, the melancholic composition apparently succeeded in touching the collective Finnish psyche and therefore in becoming a prominent part of the national audio-visual heritage.

In 1995, two main events dominate the media material. The first was that Finland joined the European Union. This was the topic of a large number of media clips, such as news and current affairs programmes, as well as sketches, and there are many background articles belonging to the EU thematic. The second event was winning the Ice Hockey World Championship, which was and has continued to be a key national experience. Ice hockey is Finland's largest sport in terms of media visibility and every spring the Ice Hockey World Championships attract a large national audience. The televised winning final in 1995 was watched by 46% of all Finns. Cultural historian Hannu Salmi has connected the boom of ice hockey to changes within the Finnish media landscape. In the 1980s, when the national team began to enjoy success, ice hockey became more visible and television rights became the subject of a struggle between public service and commercial media companies.<sup>31</sup> The victory in 1995 was the first World Championship win for any Finnish sports team and the nation was united in celebrating the victory together. In public, the victory was made into a question of national self-esteem.<sup>32</sup> Therefore, remembering the great victory also has an important role in the Living Archive.

During the third peak year of 2008, the key experiences are connected to the political elections, which received significant media attention. In Finland, the Finns Party (*perussuomalaiset*, formerly the True Finns Party) was the biggest winner in the municipal election and brought forward a strong criticism of the existing immigration policy in public discussion. Another notable political event that contributed to the 2008 peak was an event occurring outside Finland, when, in the United States, Barack Obama was elected President. However, most of the accumulated media clips during the year concerned the Kauhajoki school shooting, when a student shot and killed 11 people, including himself. After an earlier Finnish school shooting in Jokela the previous year, it was not possible to treat the Kauhajoki shooting as an individual incident, and journalists thus tried to find political and social explanations for the tragedy (for example, related to gun legislation).<sup>33</sup> The Living Archive documented this debate.

In addition to these events, many other 2008 media clips focused on the 30th anniversary of the Finnish rock festival *Provinssirock*. Even if many key experiences are connected with Finnish or world history, the audio-visual nature of the archive places a particular emphasis on audio-visual events. These events emphasise the particular role of Yle—for example, regarding television series, the Eurovision Song Contest, sports and footage of music festivals. Therefore, the national audio-visual heritage is Yle's particular heritage and its particular contribution to Finnish history; Yle owns a large proportion of the Finnish audio-visual heritage. Commercial media companies, like MTV founded in 1957, have no similar archives and there are no programmes left from the early decades of the independent Finland. During the early years of television, programmes reached large audiences, allowing programmes to create iconic images and become part of the national public memory.<sup>34</sup> However, in the

Living Archive, commercial television and radio are left aside, rendering it an important resource for the public service company Yle in the struggle among the various media to legitimise its cultural position.

### Finally: The Limits and Possibilities of Interpretation

Digital archives are organic entities that grow and change their shape as new materials are added.<sup>35</sup> The strategy of Yle's Living Archive is to grow constantly, which produces a cyclical nature for the production of this national audio-visual heritage. Annually recurring events related to the audio-visual culture have generated the most archival material over the years. In this way, Yle has created a national audio-visual annual calendar to commemorate significant moments, like Independence Day. Another strategy of the Living Archive, tackling contemporary topical issues, in turn, binds the past to this present moment. The archive actively follows current topics and events. This strategy equals Derek Kompare's notion that the legitimization of television, as in the case of archive, was based not on the existing canon, as it had been with film, but on the idea of heritage. While the canon tends to be separated from everyday life and located in a refined, timeless sanctuary, a heritage is part of the lived, historical experience of a culture.<sup>36</sup> Both of Yle's strategies have succeeded in producing the most popular content in the archive, such as responses to topical issues and death notifications for public figures.<sup>37</sup>

It is important to note that the choices of which media clips are published in the Living Archive and their output are political choices that remain invisible to archive users.<sup>38</sup> As Pajala has pointed out, the archive has limited possibilities to publish anything not shown by public service television. As a result, former legislation and cultural norms continue to restrict today's debate on the subject of audio-visual archives.<sup>39</sup> In addition, the copyright agreements significantly shape the publication choices, as not all of Yle's archive material can be published. During the first years, the only materials published were Yle's own journalistic programmes and films redeemed from old film companies. Another restriction is that all released music has to be reported to copyright organisations, but since 2015, a separate agreement has made it possible to publish many old music programmes, such as the Finnish qualification competitions for the Eurovision Song Contest and the music programme *Hittimittari* from the 1980s. A drama agreement in 2016 has also enabled Yle to release drama series and films, with long publishing rights. However, this is not only about a question of choice, since not all of the old material is actually available. Only since 1984 has all self-produced TV material been archived. In addition, the old programmes may have very poor metadata, making them more difficult to locate.<sup>40</sup>

In addition to the choice of data, we also have to take into account FiNER's specific internal limitations, which are similar to those of other rule-based

NER systems: any shortcomings in rule formulation or gazetteers result in false positives and misclassification of matched names, which if no rule applies, can make a name go altogether unnoticed. Single-word names of organisations and events are particularly difficult to identify without context clues or structural hints. This is particularly true for short texts such as the metadata entries, where the reader is often assumed to be familiar with names and their referents. In the case of rule-based NER systems, any misses and tagging errors are human in nature as they arguably reflect the system developer's ability or inability to formulate exhaustive rules, as well as their oversights when building the gazetteers.

Metadata as a research material also raises wider ontological questions of historical knowledge: What can we really know about the audio-visual cultural heritage on the grounds of metadata and FiNER analysis? Digital history researcher Michael J. Kramer discusses the relationship between historians and digital archives and describes history as a meta-metadata.<sup>41</sup> There are so many layers of interpretation: metadata is not merely descriptive, but is also already an interpretation of an archivist. Another issue stems from the fact that the results of the FiNER analysis are the lists of frequencies of names. These frequencies are not directly related to media material: FiNER as a tool has its own limitations and rules on how it interprets and categorises the metadata that has been pre-processed by the researcher, something which also brings its own specific limitations. Finally, the historical researcher interprets the results of FiNER. We suggest that it is actually already a meta-meta-metadata of the target of study. In order to understand all these layers of interpretation, we need collaboration between archivists and historians to make visible the guidelines and ways of writing metadata. The most fruitful approach would be for them to cooperate in negotiating how metadata best serve both parties.

A historian also needs to be aware of the functional logic of the digital tool for being able to recognise the bumps on the road of interpretation.<sup>42</sup> Our NER-based analysis revealed an interesting emphasis on people, events and years in the audio-visual heritage constructed by the Living Archive. However, the FiNER analysis did not shed light on why and how these certain topics are represented in the archive. Instead, what the analysis did show is why audio-visual heritage is constructed in a certain way.

Metadata, like Yle's metadata, is often messy and requires a significant amount of selection and pre-processing. In addition, the metadata material, as well as the digital tool, has limitations, as described above.<sup>43</sup> The large volume of data can compensate for some of the technical limitations. However, it is important to acknowledge that, in the humanities, the shift from small smart data to big data is not just technological; in fact, it seems to be even more of a methodological shift. Methodologically, it means the shift from close reading to distant reading. In this paradigm, instead of reading a few selected texts, we can analyse an entire collection of relevant textual data.<sup>44</sup> However, the cultural contextualising and close reading of the themes pointed out by the results of NER-based analysis still play an important role in the analytical process. Only

after the cultural contextualising and close reading do the word lists come alive, so they can animate the role of the audio-visual heritage in the construction of the Finnish imagined national community broadcast on radio and television.

### Acknowledgements

We would like to thank the research project From Roadmap to Roadshow, funded by Kone Foundation, for organising the Digital History Finland workshop. Assistance provided by Yle's Living Archive was greatly appreciated. Thanks also to editors and anonymous reviewers for their suggestions and criticisms, all of which made this a better text.

### Notes

- <sup>1</sup> An earlier, Finnish language version of this research entitled *Kekkonen, Euroviisut ja Helsinki—kansallinen audiovisuaalinen perintö NER-analyysin tunnistamana* was published in the journal *Ennen ja nyt* (2/2019). This edited version is published with the permission of the journal *Ennen ja nyt*.
- <sup>2</sup> Assmann 2008: 100.
- <sup>3</sup> Kompare 2005: 103.
- <sup>4</sup> Anderson 2007/1983.
- <sup>5</sup> Kompare 2005: 102–103.
- <sup>6</sup> Yli-Ojanperä 2018.
- <sup>7</sup> Pajala 2010: 134, 142.
- <sup>8</sup> Material is in the form of a CSV file, which is 18 megabytes in size and contains 34,816 entities.
- <sup>9</sup> Elina Yli-Ojanperä's interview 23 May 2018.
- <sup>10</sup> Marrero et al. 2013: 484.
- <sup>11</sup> In this chapter, we use a work version between two publications, which, however, is very similar to the version that has been published at the end of the year 2018. See <http://urn.fi/urn:nbn:fi:lb-2018091301>.
- <sup>12</sup> Kokkinakis et al. 2014.
- <sup>13</sup> However, see Kettunen et al. 2017.
- <sup>14</sup> Karttunen 2011.
- <sup>15</sup> Elo & Kleemola 2016: 154.
- <sup>16</sup> van Hooland et al. 2015: 13.
- <sup>17</sup> Puro 2016: 23–24.
- <sup>18</sup> Lounasmeri & Sumiala 2016: 3–4.
- <sup>19</sup> On the cult of President Kekkonen and his popular cultural image, see Kallioniemi, Kärki & Mähkä 2016.
- <sup>20</sup> Tepora 2015b: 8–11.
- <sup>21</sup> Halldórsdóttir, Kinnunen & Leskelä-Kärki 2016: 19.
- <sup>22</sup> Kurvinen 2013: 68.
- <sup>23</sup> Pajala 2006: 351–368.

- <sup>24</sup> Torsti 2012: 135–155; Ahonen 2017: 10–11, 92–95.  
<sup>25</sup> Tepora 2015a.  
<sup>26</sup> Pajala 2006: 24; Pajala 2012.  
<sup>27</sup> Cf. Sumiala-Seppänen 2007: 283; Torsti 2012: 63–73. On collective memory, see Benton 2010: 1, 5.  
<sup>28</sup> Elina Yli-Ojanperä's interview 20 June 2019.  
<sup>29</sup> Helsingius 2010.  
<sup>30</sup> Yle Elävä arkisto (Living Archive) 2015.  
<sup>31</sup> Salmi 2015: 13, 32–46.  
<sup>32</sup> Kannisto 2015: 95–96.  
<sup>33</sup> Raittila et al. 2009: 112.  
<sup>34</sup> Pajala 2010: 135.  
<sup>35</sup> Jarlbrink & Snickars 2017: 1240.  
<sup>36</sup> Kompare 2005: 105.  
<sup>37</sup> Yli-Ojanperä 2018.  
<sup>38</sup> Cf. Jarlbrink & Snickars 2017: 1229.  
<sup>39</sup> Pajala 2010: 140.  
<sup>40</sup> Yli-Ojanperä 2018.  
<sup>41</sup> Kramer 2014.  
<sup>42</sup> See also Elo & Kleemola 2016: 154.  
<sup>43</sup> See also Kannisto 2016.  
<sup>44</sup> Schöch 2013.

## References

- Ahonen, S.** (2017). *Suomalaisuuden monet myyjit: Kansallinen katse historian kirjoissa*. Helsinki: Gaudeamus.
- Anderson, B.** (2007/1983). *Kuvitellut yhteisöt: Nationalismin alkuperän ja leviämisen tarkastelua*. Tampere: Vastapaino.
- Assmann, A.** (2008) Canon and archive. In A. Erll & A. Nünning (Eds.), *Cultural memory studies: an international and interdisciplinary handbook* (pp. 97–107). Berlin and New York, NY: Walter de Gruyter.
- Benton, T.** (2010). Introduction. In T. Benton (Ed.), *Understanding heritage and memory* (pp. 1–5). Manchester: Manchester University Press.
- Elo, K., & Kleemola, O.** (2016). SA-kuva-arkistoa louhimassa: Digitaaliset tutkimusmenetelmät valokuvatutkimuksen tukena. In K. Elo (Ed.), *Digitaalinen humanismi ja historiatieteet* (pp. 151–190). Historia mirabilis 12. Turku: Turun Historiallinen Yhdistys.
- Halldórsdóttir, E. H., Kinnunen, T., & Leskelä-Kärki, M.** (2016). Doing biography. In E. H. Halldórsdóttir, T. Kinnunen, M. Leskelä-Kärki & B. Possing (Eds.), *Biography, gender and history: Nordic perspectives* (pp. 7–34). Turku: k&h, kulttuurihistoria, Turun yliopisto.
- Helsingius, B.** (2010). *Mårtenson, Lasse (1934–2016)*. National Biography of Finland. Studia Biographica 4. Helsinki: Suomalaisen Kirjallisuuden Seura. Retrieved from <https://kansallisbiografia.fi/kansallisbiografia/henkilo/1095>

- Jarlbrink, J., & Snickars, P.** (2017). Cultural heritage as digital noise: nineteenth century newspapers in the digital archive. *Journal of Documentation*, 73(6), 1228–1243. DOI: <https://doi.org/10.1108/JD-09-2016-0106>
- Kallioniemi, K., Kärki, K., & Mähkä, R.** (2016). Kekkonen ja rock'n'roll. In M. Kaartinen, H. Salmi & M. Tuominen (Eds.), *Maamme—itsenäisen Suomen kulttuurihistoria* (pp. 395–414). Helsinki: Suomalaisen Kirjallisuuden Seura.
- Kannisto, M.** (2015). Sankarit kultajuhlissa. In B. Heiskanen & H. Salmi (Eds.), *Kiekkokansa* (pp. 72–101). Helsinki: Teos.
- Kannisto, M.** (2016). Uusi, ehompi, paras? Digitaaliset tekstilouhinnan työkalut televisiotutkimuksessa. *Lähikuva*, 29(1), 62–70. DOI: <https://doi.org/10.23994/lk.56578>
- Karttunen, L.** (2011). Beyond morphology: pattern matching with FST. In C. Mahlow & M. Piotrowski (Eds.), *Systems and frameworks for computational morphology, volume 100 of communications in computer and information science* (pp. 1–13). Berlin and Heidelberg: Springer.
- Kettunen, K., Mäkelä, E., Ruokolainen, T., Kuokkala, J., & Löfberg, L.** (2017). Old content and modern tools: searching named entities in a Finnish OCREd historical newspaper collection 1771–1910. *Digital Humanities Quarterly*, 11(3). Retrieved from <http://www.digitalhumanities.org/dhq/vol/11/3/000333/000333.html>
- Kokkinakis, D., Niemi, J., Hardwick, S., Lindén, K., & Borin, L.** (2014). HFST-SweNER—a new NER resource for Swedish. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk & S. Piperidis (Eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik, Iceland: European Language Resources Association.
- Kompare, D.** (2005). *Rerun nation: how repeats invented American television*. London and New York, NY: Routledge.
- Kramer, M. J.** (2014). Going meta on metadata. *Journal of Digital Humanities*, 3(2). Retrieved from <http://journalofdigitalhumanities.org/3-2/going-meta-on-metadata/>
- Kurvinen, H.** (2013). 'En mä oo mies enkä nainen. Mä oon toimittaja': Sukupuoli ja suomalainen toimittajakunta 1960- ja 1970-luvulla. *Acta Universitatis Ouluensis B Humaniora* 113. Oulu: Oulun yliopisto.
- Lounasmeri, L., & Sumiala, J.** (2016). Promootiokulttuurin kuningas Kekkonen. *Lähikuva*, 29(2), 3–5. Retrieved from <https://journal.fi/lahikuva/issue/view/3995>
- Marrero, M., Urbano, J., Sánchez-Cuadrado, S., Morato, J., & Gómez-Berbís, J. M.** (2013). Named Entity Recognition: fallacies, challenges and opportunities. *Computer Standards & Interfaces* 35(5), 482–489. DOI: <https://doi.org/10.1016/j.csi.2012.09.004>
- Pajala, M.** (2006). *Erot järjestykseen! Eurovision laulukilpailu, kansallisuus ja televisiohistoria*. Nykykulttuurin tutkimuskeskuksen tutkimuksia 88. Jyväskylä: Jyväskylän yliopisto.



- Pajala, M.** (2010). Cultural memory and its limits on the Finnish public service broadcaster's online archive. *Critical Studies in Television*, 5(2), 133–145.
- Pajala, M.** (2012). Televisio kulttuurisen muistin medianä: Miten itsenäisyyspäivä alkoi merkitä sotamuistelua? In E. Railo & P. Oinonen (Eds.), *Media historiassa*, Historia mirabilis 9 (pp. 127–150). Turku: Turun historiallinen yhdistys ry.
- Puro, J.-P.** (2016). Puhumalla hallitseminen: Vahvan johtajuuden ilmentyminen Kekkosen radiopuheissa vuosina 1937–1967. *Lähikuva*, 29(2), 22–42. Retrieved from <https://journal.fi/lahikuva/issue/view/3995>
- Raittila, P., Haara, P., Kangasluoma, L., Koljonen, K., Kumpu, V., & Väliiveronen, J.** (2009). *Kauhajoen koulusurmat mediassa*. Tampereen yliopisto, tiedotusopin laitos: Journalismin tutkimusyksikkö. Retrieved from <http://tampub.uta.fi/bitstream/handle/10024/65360/978-951-44-7654-9.pdf>
- Salmi, H.** (2015). Kiekkokansa ryhmäkuvassa. In B. Heiskanen & H. Salmi (Eds.), *Kiekkokansa* (pp. 13–46). Helsinki: Teos.
- Schöch, C.** (2013). Big? Smart? Clean? Messy? Data in the humanities. *Journal of Digital Humanities*, 2(3). Retrieved from <http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>
- Sumiala-Seppänen, J.** (2007). 'Joku raja ja loppu se on saatava tällaisille keskusteluille!' Jatko aika ja kansallisten kertomusten konflikti 60-luvun suomalaisessa televisiossa. In J. Wiio (Ed.), *Television viisi vuosikymmentä: Suomalainen televisio ja sen ohjelmat 1950-luvulta digiaikaan* (pp. 280–291). Helsinki: Suomalaisen Kirjallisuuden Seura.
- Tepora, T.** (2015a). Mikä tekee 'sota-ajasta' muistettavan? Sota ja kollektiivinen muistaminen. *Ennen ja nyt*, 2015(2). Retrieved from <http://www.ennenjanyt.net/2015/08/mika-tekee-sota-ajasta-muistettavan-sota-ja-kollektiivinen-muistaminen/>
- Tepora, T.** (2015b). Sota, Tuntematon ja Mannerheim. *Futura*, 34(3), 5–11.
- Torsti, P.** (2012). *Suomalaiset ja historia*. Helsinki: Gaudeamus.
- van Hooland, S., De Wilde, M., Verborgh, R., Steiner, T., & Van de Walle, R.** (2015). Exploring entity recognition and disambiguation for cultural heritage collections. *Literary and Linguistic Computing*, 30(2), 262–279. DOI: <https://doi.org/10.1093/llc/fqt067>
- Yle Elävä arkisto** (2015). *Myrskyluodon Maija monella tavalla*. Retrieved from <https://yle.fi/aihe/artikkeli/2015/07/30/myrskyluodon-maija-monella-tavalla>
- Yli-Ojanperä, E.** (2018, helmikuu). *Yle Elävä arkisto*. Avoin digitaalinen kulttuuriperintö—course, slides. Yle.
- Yli-Ojanperä, E.** Producer-in-chief, Yle Living Archive. Interviewed by Maiju Kannisto by email 23 May 2018 and 20 June 2019.

### *Data systems*

FiNER: <http://urn.fi/urn:nbn:fi:lb-2018091301>

Tools for processing FiNER output: <https://github.com/pkauppin/finer-utilities>