

NEWSREELS VERSUS NEWSPAPERS VERSUS METADATA

A COMPARATIVE STUDY OF METADATA MODELLING THE 1930S IN ESTONIA

Indrek Ibrus
Tallinn University
Narva road 27
10120 Tallinn
Estonia
indrek.ibrus@tlu.ee

Maarja Ojamaa
Tallinn University
Narva road 27
10120 Tallinn
Estonia
maarja.ojamaa@tlu.ee

Abstract: This article offers a comparative take on the ways audiovisual versus verbal digital archives model our understanding of the past. We focus on content metadata schemas and their role in modeling histories and framing the uses of audiovisual databases. Our empirical corpus includes verbal and audiovisual objects from the five-year period just before the World War II (1935-1939) as presented in two digital databases – the Analytic Bibliography of Estonian Journalism and the Estonian Film Database. The article compares how the different metadata schemas for newspaper articles and newsreels model their objects. As a consequence, metadata schemas shape contemporary perceptions of historical realities in different ways.

Keywords: metadata, modelling systems, audiovisual archives, Estonia, 1930s, newsreels, cultural semiotics

The amount and variety of information about the 20th century that we can freely access online is in every respect remarkable. Contemporary digital databases as Europeana, Internet Archives or NYPL Collections offer anyone interested a chance to go far beyond individual textbooks and occasional faded photographs: one can juxtapose materials in different modalities and compare or contrast diverse aspects of the past from various viewpoints. This applies to both users with general or hobbyist interest in the past as well as professional historians.

The digital archival turn has evidently met a wide spectrum of responses. Wolfgang Ernst¹, for one, advocates the emergence of “a new memory culture” in which the power hierarchies that frame knowledge at classical analogue archives dissolve. Yet, Andreas Fickers²—having a more skeptical view—points out that the digital shift into the age of abundance has not been accompanied by a similar shift of source criticism, indicating the scarcity of adequate

¹ Wolfgang Ernst, *The Archive as Metaphor: From Archival Space to Archival Time*, 2004, <https://www.onlineopen.org/the-archive-as-metaphor>

² Andreas Fickers, 'Towards A New Digital Historicism?: Doing History In The Age Of Abundance,' *VIEW Journal of European Television History and Culture*, 1, 1, 2012, 12-18.

contextualizing metadata. Indeed, the ways that archival materials can be discovered from the penumbra of increasingly vast databases and then interpreted and reused to make sense of our past are largely determined by the attached metadata. What the present article seeks to discuss are the ways that distinctions between verbal and audiovisual representations of the past appear at the level of descriptive metadata.

We will first introduce our theoretical framework, which originates mainly from the works of Russian–Estonian semiotician Juri Lotman (1922–1993), the founder of the Tartu-Moscow school of cultural semiotics. Lotman's explanations of culture and its operating mechanisms in terms of translational communication between different sign systems or cultural languages allows for discussing metadata as a certain language or, more exactly, a metalanguage for culture. This implies that the limits and affordances of metadata shape what can—and cannot—be expressed about data, i.e. which aspects of an object can be rendered meaningful and which are conditioned to be discarded. This in turn also frames our understanding of the past itself that the data mediates. The theoretical part of the paper will be followed by an analysis of the metadata schemas of two digital databases: the Estonian Film Database and the Analytic Bibliography of Estonian Journalism, concentrating on and containing newsreel and newspaper materials from the years between 1935 and 1939.

1 Cultural Semiotic View on Metadata

Let us first discuss the form and functions of metadata. Following Jeffrey Pomerantz's definition metadata is “a means by which the complexity of an object is represented in a simpler form”³. The metadata profile of a content entity in a database represents the entity through brief statements in a more concise and ordered manner (than in reality). The entity can thereby be rendered searchable among and relatable to other entities within a given collection. Via metadata statements and schemas, descriptive dominants of objects are chosen in accordance with the overall focus and interests of a database. A newsreel, for example, can be represented both as a film and as an historical document or as an artefact to be preserved or as a resource to be reused and, accordingly, different aspects would be outlined in the database entry. For example, adding into the database vocabulary a keyword describing the mode of filmic expression (e.g. “aerial shot”) directs the users to recognise and take notice of this modality of the given film. However, it will also institute a shift from a secondary descriptive status to the core of the audiovisual form and mediation of the past within the database—and thus influence the future usage of the database.

Continuing on the path that Pomerantz lays out, we might further specify metadata schema as “a very simple language”⁴. Its structure and range of functions is boiled down to communicating invariant information about a resource, which means that metadata schemas as artificial languages are much less flexible and ambiguous compared to natural human language, which serves functions besides communicating information. However, as long as we agree that a metadata schema is an “ordered system which serves as a means of communication and employs signs”⁵, it is compatible with the cultural semiotic definition of language and we can discover further characteristics of metadata through this prism. Most importantly, from a cultural semiotic perspective a language not only simplifies the complexity of the surrounding reality, but models it, i.e. it functions as a modelling system. Lotman explains the latter as follows: “a structure of elements and rules of their combination, existing in a state of fixed analogy to the whole sphere of the object of perception, cognition, or organization”⁶. In another text he claims briefly: “the modelling influence of the metalanguage on its object is inevitable”⁷.

In the case of archived objects we should hence distinguish between two modelling operations: the newsreels and newspaper articles both contain elements and rules for combining them even though in the first case these elements

3 Jeffrey Pomerantz, *Metadata*, MIT Press, 2015, p. 11.

4 Ibid., p. 30.

5 Juri Lotman, *The Structure of the Artistic Text*, University of Michigan, Department of Slavic Languages and Literatures, 1977, p. 6.

6 Juri Lotman, ‘The Place of Art among Other Modelling Systems,’ *Sign Systems Studies*, 39, 2/4, 2011 [1967], <http://www.sss.ut.ee/index.php/sss/article/view/SSS.2011.39.2-4.10>

7 Juri Lotman, *The Structure of the Artistic Text*, University of Michigan, Department of Slavic Languages and Literatures, 1977, p. 21.

belong to a visual sign system, and in the second case into a verbal one. In short, both newsreels and newspapers (as any other text) create a model of the world that they mediate. The second modelling operation takes place when the elements of these texts are set into the state of fixed analogy with the elements of the metadata schema, which thereby model the archived texts, but consequently also the world that these mediate.

Hence, in relation to the archival object, a description of it (e.g. a metadata profile) appears as a metatext and in relation to the cultural language in which the object is mediated (e.g. audiovisual or verbal), metadata schema appears as a metalanguage. What happens in the process of metadating is thus a translation from one language to another, while each of the languages has its own specific system of rules for encoding, i.e., for establishing semantic equivalence. While a newsreel story translates an historical event into the audiovisual language manifested in cinematic imagery, the metadata profile translates the newsreel story and its historical events into a verbal language of keywords. Thus, both a newsreel story and its metadata profile model an historical event in accordance with the affordances and constraints of their operating sign system. This form of sequencing can be described as a process of intersemiotic translation in terms of Roman Jakobson.⁸ Given that there are no exact equivalents between the meaningful elements (i.e. verbal and audiovisual signs) of the languages involved, but only a “conventional system of equivalences”⁹, these intersemiotic translations are both inexact and creative. By “creative” we do not mean artistic, but simply that results of the process are not automatic.

This suggests that metadata ceases to be a mere road sign to the content entity, it becomes a model of the content entity—a model that is not automatic and is, therefore, at every instance a specific attempt to stand for chosen characteristics in the corpus of object texts. Metadata models therefore frames the interpretational frameworks for texts in the database as well as guides other uses of these texts and the creation of new texts. Relatedly, changes in the metadata schema potentially also bring along transformations in the usage scenarios of the database and its objects. That is, while the relations of conditional equivalence between the elements of a complex newsreel and the elements of its simplified metadata profile are created, the model created in turn shapes contemporary perceptions of the past that the newsreel mediates, including present and future uses of the newsreel.

While metadata models are designed to model a corpus of object-texts, they are also designed to serve a specific range of search queries or use scenarios. An observational position can therefore be taken that reads them also as modelling these possible scenarios and uses. As such, the forms of metadata could be understood as boundary objects following Susan Star and James Griesemer,¹⁰ or as cultural boundaries in terms of Lotman himself¹¹—part of different cultural sub-systems, being shaped by, as well as shaping, hence becoming a translatory language connecting various domains. Our argument can therefore be framed by the more general context of research into database culture and into the ways that (media) historical research increasingly depends on digital databases^{12,13}. At the same time, studies into the dimensions of metadata that we hereby term as modelling, are relatively scarce. Among the few are Johanna Drucker's¹⁴ more general take on the combination of descriptive and performative powers of metadata's content modelling activity, Jeremy W. Morris' explications on how “metadata act as a new nodal point for gatekeeping the music consumption process”¹⁵, or Dagmar Brunow's¹⁶

8 Roman Jakobson, 'On Linguistic Aspects of Translation,' *Selected Writings. 2. Word and Language*, Mouton, 1971, 260-266.

9 Juri Lotman, *Universe of the Mind: A Semiotic Theory of Culture*, Indiana University Press, 1990, p. 15.

10 Susan Star and James Griesemer, 'Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39,' *Social Studies of Science*, 19, 3, 1989, 387-420.

11 Juri Lotman, 'On the Semiosphere,' *Sign Systems Studies*, 33, 1, 2005, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.693.9961&rep=rep1&type=pdf>

12 e.g. Kate Theimer, Joshua Sternfeld, Katharina Hering and Michael J. Kramer, 'Digital Historiography and the Archives,' *Journal of Digital Humanities*, 3, 2. See also the whole issue.

13 e.g. Wendy Hui Kyong Chun, 'The Enduring Ephemeral, or the Future Is a Memory,' *Critical Inquiry*, 35, 1, 2008, 148-171.

14 Johanna Drucker, 'Text Models: Performative Metatexts, Metadata, and Markup,' *European Journal of English Studies*, 2, 2007, 177-191.

15 Jeremy W. Morris, 'Making music behave: Metadata and the digital music commodity,' *New Media & Society*, 14, 5, 2012, 850-866.

16 Dagmar Brunow, 'Curating Access to Audiovisual Heritage: Cultural Memory and Diversity in European Film Archives,' *IMAGE [&] NARRATIVE*, 18, 1, 2017, 97-110.

study on how metadata sources the paratexts that contextualize the archival material, while her more specific interest is the topic of diversity of the possible contextualizations.

The cultural semiotic understanding of culture as a polyglot, i.e. communicating simultaneously in different languages, consequently means the impossibility of a single perfect metalanguage to describe cultural complexity in a comprehensive manner. However, the strive for creating self-descriptions can be regarded as a cultural universal¹⁷. The concept of self-description, borrowed from cybernetics, has become central for understanding the mechanisms of cultural evolution. The latter departs from an active dialogue between the ideal unity established in self-description and “the irregularity of a real semiotic map”¹⁸. As the creation of self-descriptions is simultaneously a creation of (meta)languages for their implementation, it means that cultural object languages (e.g. cinematic, journalistic) exist in a state of constant tension with descriptive languages, because of the slowing down effect that descriptions have over any system in culture—the dynamic evolution halts or slows down when described, codified, regulated or standardised. While the dominance of object languages leads to cultural schizophrenia, to the collapse/disintegration of unified identity, the dominance of metalanguages refers to stagnation.¹⁹

If we implement this viewpoint in the context of audiovisual history, then the inability to find a metalanguage for describing different films in a common metalanguage as part of a unified corpus would mean that film ceases to form a coherent part of a given culture’s identity. This is because of the lack of means for conceptualising and describing the invariant and shared elements of films created in this culture. Had there existed a rigidly standardized metadata schema, able to describe all existing films, this would have meant that nothing new was created within this system and it ceased to evolve. Thus, there are two opposing tendencies that need to be taken into account while aggregating texts of different media onto common cultural heritage platforms. The first one is the creation of interoperability of metadata for facilitating the perception of a kind of higher level cultural whole in ways relevant to the given user. The parallel tendency, striving towards the opposite direction, is preservation of individuality of metalanguages to maintain the untranslatable areas, where the creation of new meanings is based.

Against this background, we propose that the collections of newsreels and of print newspapers consist of two distinct ways of memorizing—a distinction that is evident already at the level of metadata. The recognition of their relation of complementarity (not substitution) becomes especially important from the viewpoint of memory curation, and in educational contexts. The current efforts to raise media literacy among European youth testify to the need to realize something that tends to remain unnoticed—the ways media model their content in accordance to their affordances and constraints. At the same time it is also important to remember that meanings and memory are not located inside media objects, e.g. in the footage, but in their usage, remediation, recontextualization²⁰. And the latter is largely determined by (verbal) metadata.²¹

2 Databases and Corpora

Our empirical corpus includes data (verbal and audiovisual objects) and their metadata from the five year period between 1935 and 1939, just before the launch of the Second World War as presented in two digital

17 Peeter Torop, ‘Culture and Translation,’ *Applied Semiotics / Sémiotique appliquée*, 9, 24, 2010, 11–17.

18 Juri Lotman, ‘On the Semiosphere,’ *Sign Systems Studies*, 33, 1, 2005, <http://citeseeerx.ist.psu.edu/viewdoc/download?doi=10.1.1.693.9961&rep=rep1&type=pdf>

19 Егоров, Б.Ф., М.Б. Игнатьев, and Ю.М. Лотман, ‘Искусственный интеллект и метамеханизм культуры,’ *Russian Studies*, 4, 1995, 277–287.

20 Jaimie Baron, *The Archive Effect: Found Footage and the Audiovisual Experience of History*, Routledge, 2014.

21 Deb Verhoeven, ‘As Luck Would Have It: Serendipity and Solace in Digital Research Infrastructure,’ *Feminist Media Histories*, 2, 1, 2016, 7–28.

databases—the Analytic Bibliography of Estonian Journalism (BIBIS)²² and the Estonian Film Database (EFDb).²³ The first corpus includes all the articles from the newspaper *Postimees*—historically the first, and today still the largest, non-tabloid newspaper in Estonia. The second corpus consists of all indices of these articles in the BIBIS database. The third corpus includes all Estonian newsreel stories of the same period. Finally, the fourth corpus consists of all indices relating to all the newsreel stories in that database.

Table 1. *Quantitative differences in the corpora*

Database	Newsreels/articles 1935–1939	Keywords
Analytic Bibliography of Estonian Journalism	103,835	33,014
Estonian Film Database	785	8,508

Table 1 indicates significant differences in quantity. *Postimees* was by the second half of the 1930s a well-established daily paper, with southern Estonia and especially the Tartu county as its main domain. It therefore published thousands of news stories each year on everyday life and business in the Tartu area and elsewhere in Estonia—from the political news to the most practical information relevant to the people of Tartu. It needs to be noted, however, that in 1934 a political coup took place in Estonia. Konstantin Päts, head of state at the time, consolidated most of the executive and legislative power into his own hands and, among other initiatives, took control of all the newspapers, including *Postimees*. Therefore, when it comes to political, governance or international affairs type of news they were all controlled by the propaganda office in Tallinn. Little initiative was left to newspapers themselves, and the period between 1934 and 1940 is known as the “silent era” in Estonian history.

Regarding newsreels, their production in the 1920s and early 1930s was rather minimal—very few newsreels were produced each year between 1921 and 1933. By the early 1930s, film production was suffering from economic hardship. As a reaction and as a result from the ‘silent era’ governance practices, a new Cinema Law was drawn up and adopted in 1935. This law obligated, among other things, all cinemas—which were strictly controlled by the state and municipalities—to show newsreels as part of every screening. In parallel, the Kultuurfilm studio, which had been producing most of the films including newsreels was nationalised. Therefore, both the production and distribution of newsreels that started in 1935 was controlled and facilitated by the state. In terms of production quantities, however, an exponential growth in production occurred, culminating in 1937 with 262 newsreel films. Afterwards production stabilized at a level of slightly below 150 films a year.

Production of newsreels in Estonia 1921-1940

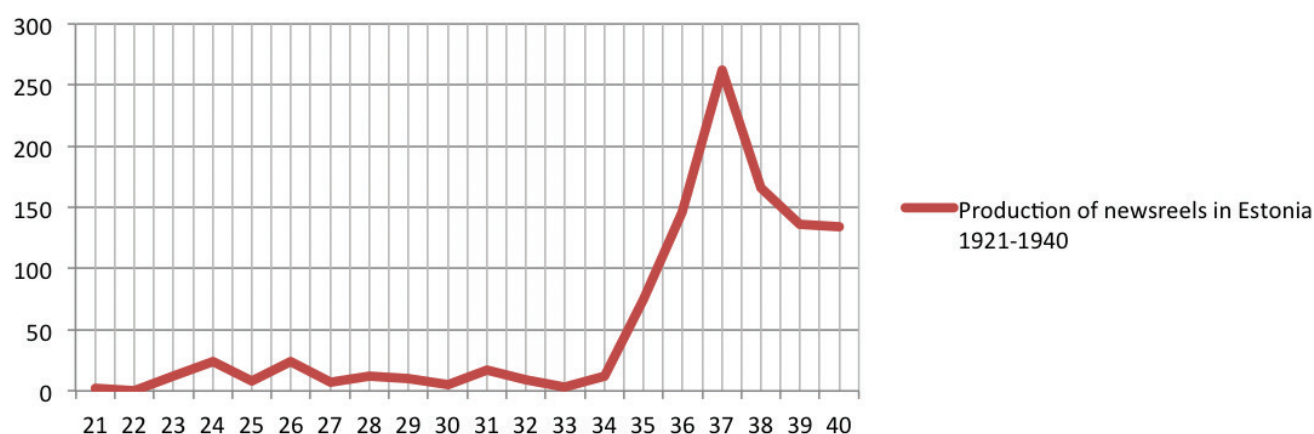


Table 2. The evolution of newsreels production in Estonia in the pre-war period.

²² <http://galerii.kirmus.ee/biblioserver/>

²³ <http://www.efis.ee/>

A comparison between the two digital databases—the Analytic Bibliography of Estonian Journalism and the Estonian Film Database—accentuates a significant difference in genre variety. While in the latter, EFDB, genre is relatively uniform (newsreel films), the newspaper *Postimees* naturally varies considerably: the most recurring metadata description is, for example, Message/News [Sõnum], which had no audiovisual equivalent. The two databases also have significant differences. The BIBIS database is produced and managed by the archival library of the Estonian Literary Museum. The library was created already in 1909 and has been working on BIBIS as its online bibliography for periodicals since 2004. EFDB was launched as a NGO, co-founded in 2009 by all the main stakeholders in Estonian audiovisual culture and heritage governance. Even by international standards, EFDB is a significant initiative in Europe as it is effectively a digital and multimodal filmography of Estonia with the aim to develop elaborate meta-descriptions for all Estonian films produced. Descriptions do not only include the standardised indices for all films, but other written texts and peripheral materials about these films (sketches for costume designs, making of photos, posters, memoirs, reviews etcetera—descriptions that are modally very rich.)



Figure 1. Estonian Film Database with Advanced Search subsite in English.

The approaches of the two databases to indexing are rather different. As EFDB was to have a rich set of data and their service was being explicitly directed to a broad audience of online users, they saw that the existing metadata standards and schemas were not satisfactory for them. For this purpose they have created their own metadata-schema.²⁴ The focus is on describing each individual film and its paraphernalia in maximum detail, departing in the first place from the particularities of each film. In comparison, the vocabulary of BIBIS is strictly pre-structured.

²⁴ Interview with EFDB editor-in-chief Hagi Šein on 24 November 2017.

As an established memory institution with an official preservation mandate—and with the main purpose to serve the researcher community—they utilise the Estonian version of the Universal Decimal Classification (UDC) as their main indexing method. Relatedly, the occurrence of a corpus is also clearer in BIBIS compared to EFDdb, where the focus is more put on individual films. More than 70 percent of keywords are, for example, used only once. In comparison, the figure for BIBIS is 43 percent. In BIBIS the annotator determines only the UDC index, based on which set of keywords that is automatically generated. Therefore, there are several cases where the profile of a text includes a keyword which does not correspond to the actual content of the article.

In contrast, the EFDdb profiles depend more on the annotator's subjective gaze and the whole metadating process is exploratory. Currently the EFDdb administrators are also looking into ways for employing user-generated metadata in addition to the work of hired professional annotators. This indicates one of the main differences between the two databases: how they model their users. While EFDdb aims to address a more ambiguous audience, BIBIS is mainly targeted at researchers. This is a key factor in determining the metadata schema and vocabulary for a particular database as different schemas serve different purposes. While researchers might prefer “complex granular search availability”,²⁵ other audiences are more open to and attracted by experimental and curated ways of discovering content, including image-based²⁶ and event-based²⁷ retrieval. On the level of the user interface, BIBIS offers four broad categories for browsing keywords, in addition to a general search function: institutions, persons, place names, and other keywords. The equivalent categorization of vocabulary on EFDdb is way more refined and includes: content keywords, items/details (iconographic), keywords of the film field, film tonality, UDC, topics (feature films), topics (documentaries), events, place names, buildings and objects, institutions, persons shown, persons talked about, persons talking, dates (of the events depicted), exterior locations, interior locations, collection name, fund name.

Given that the material of expression, i.e. the sign types of the initial corpuses are different—moving images in one case and words in the other—we should also ask how this affects the process of providing texts with verbal keywords. One possible way to approach the issue is located on the level of sign types themselves: words function as symbols and have an arbitrary relationship to their signified, whereas images signify via the relationship of iconicity. In the filmic context, however, the latter can be debated, because mediating one and the same object from different angles or shot lengths can attribute different, and even contrary meanings. Therefore, filmic mediation also has a definitive symbolic dimension.

Moving from the level of signs to the level of texts, they seems to have more to do with combinations of signs than with individual sign types. Social semiotician Gunther Kress has claimed that “the world told is a different world to the world shown,”²⁸ and proposed that the crucial difference lies in the organisational logics of telling/writing and showing. Verbal texts are based on the temporal logics of sequencing elements in time, whereas images are governed by the spatial logics and simultaneity. As the examples below specify, such distinctions can be noticed in the ways that newsreels and newspaper articles mediate the world. While written articles proceed in a word by word sequence following syntactic rules, newsreels are destined to reveal what happens to be within the frame simultaneously with the object and/or subject in focus. Therefore, the set of keywords attributed to newsreels is usually richer in elements that can be visually perceived, but might not be conceptually related to the topic of the text.

Within our corpus, the keyword sets for verbal articles usually favour abstract categories and relate the object/subject in focus to other general fields of human knowledge. This is supported by the logics of the Universal Decimal Classification itself, which functions primarily as a system for ordering and contextualising textual wholes on the basis of their overall subject, while the indexing system employed at EFDdb supports reordering and recontextualising texts on the basis of individual elements or layers of the text. From this viewpoint both systems are concentrated on the content level,

25 Interview with British Film Institute's Head of Data, Stephen McConnachie on 19 October 2017.

26 T. Tommasi, et al., 'Beyond Metadata: Searching Your Archive Based on Its Audio-Visual Content,' *Proceedings of 2014 International Broadcasting Convention*, Institution of Engineering and Technology, 2014.

27 Chiel van den Akker, et al., 'Historical Event-based Access to Museum Collections,' *Applied Artificial Intelligence*, 25, 2010.

28 Gunther Kress, *Literacy in the New Media Age*, Routledge, 2003, p. 1.

i.e. on *what* has been mediated (instead of *how* it has been done) and are thus more or less applicable irrespective of the medium of the data. In addition, it is clear that newsreels and newspapers do not exist in cultural vacuums and texts expressed in visual sign systems are in complex ways influenced by verbal sign systems and vice versa. In addition, the process of reading newspaper articles also evokes mental images motivated by the reader's visual cultural memory. In the cultural semiotic perspective newsreels and newspapers, thus, model one and the same world, but do it in their medium-specific ways, and this why the modelling operations often results in completely different models.

3 How and What the Metadata Schemas Model

When grouping the EFDb vocabulary into four, analogously to BIBIS, it appears that the share of place names is equal (both ~ 3–4%), while the shares of persons (~ 9% vs ~ 36% in BIBIS) and institutions (~5% vs ~13% in BIBIS) is lower on EFDb, and the share of other keywords (~ 83% vs ~48% in BIBIS) is higher. While the constant value of place names needs further discussion, the inequalities in the other three could be seen as related to the higher number of categories in EFDb. While this indeed poses some difficulties for the comparability of the databases, several interesting differences and overlappings can still be deduced from looking at the frequencies and hierarchies of keywords in each.



Figure 2. Word cloud of top 75 most used keywords in EFDb for newsreels between 1935 and 1939 (references to years and decades excluded).



Instead, in BIBIS the category of location dominates over that of time. As newspapers are self-evidently tied to their contemporaneity, their modality does not lend itself easily for generalising representations on decades or other extended periods. Newspapers do not chronicle time, they acquire this function later, when archived and organised for reuse. Instead, their function is to inform readers about the newest developments across their chosen target territories. It may be suggested that for geographical reasons place names exist as a distinct searchable and browsable category. Tartu as the location of publishing *Postimees* is obviously at the top of the list (12,672), followed by Tallinn (3,043), and (after other keywords) by Elva (a small resort town close to Tartu). Soon after, other Estonian cities follow—quite

systematically from the larger ones (Pärnu, Narva) to smaller ones. Additionally, at the top of the BIBIS keyword list are the Universal Decimal Classification terms referring to local municipal administration at specific, respectively urban (i.e. mostly in Tartu, 2,412) and rural (locations 2,279), locations. This models, on the one hand, the function of *Postimees* during the “silent era” to cover extensively local government, but not to write about national politics and governance. But it also refers to the modelling process of the database, as these news stories denote concrete processes in localities/municipalities which enables the database to serve extensively the search interests tied to places.

In parallel, in EFDdb, the locations tend to be much more specific—locations such as the Port of Tallinn, Tallinn Art Hall, Kadriorg Castle (presidential palace and residence), Tallinn Railway Station, Kadriorg Stadium dominating. These are the places where some significant activities, often ceremonial, tended to happen and were recorded for newsreels; people arriving or leaving, competitions, exhibitions or political ceremonies taking place. While all most depicted sites are located in Tallinn, Estonia’s capital, the place category Tallinn is not among the top 50 keywords. It could be that the city “around” the depicted site was deemed irrelevant or too general for annotators. Or perhaps, it was presumed to be essential knowledge for modeled audience (i.e. Estonian internet users). Interestingly, the city of Tartu was tagged twice as much as Tallinn. Perhaps it was the case that to the annotators (living in Tallinn) Tartu and its visual premises appeared as notable “other”, a more distinct entity to be referenced, and possibly indicating how cultural circumstances of indexing might shape eventual modelling work of databases and their metadata schemas.

In terms of object-text corpuses similar differences to the above can be recognised across different categories. In the case of EFDdb visually attractive spheres are prioritized: politics is dominated by different events at national level (mostly featuring the then head of state/president (the status changed) Päts or head of military, general Johan Laidoner and by visiting delegations (e.g. arrivals) at the international level. It is notable also how the state-owned Kultuurfilm extensively covered the (likewise visually attractive) defence forces, a topic that hardly features at all in *Postimees*, the newspaper. Politics is dominated in *Postimees* by local affairs and is represented by reports on decisions by, and changes in, the administration affecting readers’ everyday lives as opposed to special events and ceremonies.

Representations of sports are high in both lists—this news genre was by then already well established in all media and both databases model that reality. But again it is noticeable how in newsreel stories and relatedly in EFDdb indices the visually as well as politically attractive motosports, heavy lifting and military sports are positioned higher. In *Postimees* reports on popular ballgames (football, handball, basketball, volleyball) dominate instead. EFDdb/newsreels also become distinct for frequent representation of youth sport and other extra-curricular events and activities. That is, education and youth organizations are represented in newsreels mostly by events, ceremonies and glimpses at youth camp activities, a possible characteristic of the 1930s and the authoritarian era.

The apparent rationale to represent military, authorities, sports and visually attractive logistics seems to have lead to an anecdotal case that, according to the database, one of the most featured persons in the newsreels was major general Otto Sternbeck (1884–1941), then minister of roads. He was also head of the Estonian Shooting Association, at the time when the Estonian national team won two consecutive World Championship titles. While he is a relatively unknown figure from the contemporary perspective (compared to other leading politicians of the time) his openings of new roads, leaving and arriving with the national shooting team and attending celebrations, made him very visible in newsreels.

What could be understood as characteristic to medium specificity of representations in news is the way in which arts appear in both frequency lists. In EFDdb visual art comes first as, presumably, a form of arts that could be well represented in silent newsreels of the time. Openings of larger exhibitions in Tallinn’s then nascent Art Hall were probably visually attractive too. In BIBIS, however, arts first appear through eight interrelated UDK sub-categories referring to buildings of various kinds. These stories are mostly about new buildings across Estonia, their planning and construction processes, their openings. In parallel, in the EFDdb list an explicitly medium-specific category modelling the potential search interest—“Rural and urban sightseeings”—featuring buildings across the country,

is high. In BIBIS, at the same time, among arts comes in second place a set of UDC categories related to books and book-related events, categories modally more suitable for newspapers than film. This assumption is supported by well-known Estonian writers of the time (Anton Hansen Tammsaare, Oskar Luts) being represented nearly as often as leading state figures.

In general, however, EFDdb includes more markers of medium specificity. The form or modality of newspaper stories are marked only in the case of more exceptional genres—feuilletons, causeries, satirical pieces, essays, reportages. In comparison, EFDdb makes a much stronger effort to model the modalities of newsreels and other films. They have added a vast set of film-specific indices, special vocabulary for film tonality, two separate vocabularies for film topics (one for fiction and another for documentaries), open vocabularies for all buildings, sites, and events represented. And there are medium-specific distinctions for people represented (people shown, people talking, people talked about). The active use of categories such as actors or activities could also be understood to derive from the need to model audiovisual narratives. That is, while the newspaper database appears to privilege nouns, the corpus of newsreels metadata is inclined towards verbs and suffixes indicating activities. In other words, audiovisual memory as modeled by metadata that stress concrete processes rather than more abstract phenomena, which is the case with verbal journalistic memory.

The upper sections of the EFDdb frequency list demonstrate the efforts to mark what is depicted—with concrete buildings, sites or people emerging at the top. But we also see in the top distinctive medium-specific markers of audiovisual modalities —“sightseeings”, “observational films”, “panoramas”, “nature views”. The potential richness of audiovisual forms and the apparent desire to model these varieties with metadata could also be exemplified with indices in the lower section of the frequency list—categories such as “spectacular” and “anxious” (both in terms of film’s tonality, whereas the first is additionally marked as neutral, the second as negative), “interview film”, “parallel montage”, “aerial shots”, “dance scenes”, “nature film”.

What these examples suggest is that the schemas and vocabularies of the EFDdb not only model the audiovisual objects in their particularities, but also the potential search interests or the cultural contexts of internet users/ searchers. Examples of “parallel montage” in the 1930s could be of interest to film historians, “interview film” to film professionals potentially searching for footage, “panoramas” or “aerial shots” could be relevant for historians, “nature films” for biology teachers or, indeed, the general public. As such these vocabularies have emerged as a “language” modelling both cultural domains—the films (and what these depict) and their uses in the imagined futures.

In comparison, our study also demonstrates how the UDC system of BIBIS is effectively path dependent on the functions of an archival library—to categorise all holdings in the library in order to manage their preservation systematically. The UDC system might be too general and misleading for online queries. That is, it cannot be used effectively to model the textual elements in their particularities nor to model potential contemporary search interests. Take, for instance, a set of categories relatively near the top in *Postimees*’ frequency list—electric energy, electrification, lighting, lighting equipment. All were used 557 times for our five year period—meaning that they always come together as UDC sub-categories. Thus, if a database user is, for instance, interested in articles on lamps she/he will mostly get articles on national strategies for electrification or building of new power plants. Therefore the metadata schema may easily mislead.

4 Newsreels and News Articles—Differences in Coverage

We propose to study three random examples of differences in the mediation of historical events. Our first case is the signing of the Estonian constitution by Konstantin Päts in 1937. In EFDdb this is represented by a 24-second newsreel story *President K. Päts Signs the New Constitution on August 17, 1937*.



Figure 4. Still from the newsreel story President K. Päts Signs the New Constitution on August 17, 1937.

In BIBIS this event is covered by two articles both specified as an Overview. The entry on the first article is minimal and all of its keywords are exhausted in the more elaborate one of the second titled *State Holder on the Enforcement of the New Constitution*, which is essentially a reprint of the president's speech on the occasion. The index of the EFDb newsreel story is richer in keywords. But it has to be taken into account that when it comes to people, institutions and locations, they are mentioned on BIBIS only in case they are directly involved in the event, whereas in newsreels, for example, people might half-accidentally appear in a scene and thereby get to be mentioned by name on the metadata profile. In the case of BIBIS, the institution is mentioned rather than the name of the person, which is represented only if it appears in the title or the annotation. So we have a list of nine persons in EFDb for this event, while in BIBIS there's just president Päts under the annotation and a list of institutions under the keywords. A similar specificity pertains to the location of the event, which is not mentioned in BIBIS as well as to the items/details appearing on the film ("letterpress", "vintage fireplace"), but not directly related to the phenomenon of constitution signing.



Figure 5. Stills from the newsreel story Estonian Art Exhibition in Budapest (1939).

Our second example is an exhibition of Estonian art in Budapest in 1939. In EFDdb this is represented by a 53-second newsreel story *Estonian Art Exhibition in Budapest* and by the message *The Opening of Estonian Art Exhibition in Budapest* in BIBIS. This event appears as predominantly a political one in BIBIS—but as a cultural one in EFDdb. In BIBIS the modelling process is contributed to both by UDC and additional keywords, which include “foreign policy”, “border conflicts”, “interstate negotiation”, and “embassies”. The EFDdb profile, at the same time, refers mostly to art-related aspects (including artist names) and only briefly to the political dimension using the keywords “Estonian–Hungarian relations”, “state characters” and to Hungarian statesman Miklós Horthy as well as to Estonian minister of foreign affairs Karl Selter. It needs to be emphasized that neither the article nor the newsreel story mentions political negotiations, border conflicts or ideological issues.



Figure 6. Still from the newsreel story *Phenomenal Accident at the Port of Tallinn* (1939).

Our third example evolves around the sinking of a steamship at the port of Tallinn in 1939. This event is represented by a 86-second newsreel story *Phenomenal Accident at the Port of Tallinn* in EFDdb and by the message *A Steamship Drowned at the Port of Tallinn* in BIBIS. The latter is another example of a misleading effect caused by the automatic generation of keywords based on the UDC index. Namely, from the keyword “helping the shipwrecked”, especially in Estonian [“*merehädaliste abistamine*”], the user could get an impression that the steamship was manned, while in fact it was empty. Both of the texts concentrate on the rescuing works, and the lifting of the ship out of the water. As the EFDdb profile includes general keywords such as the above-mentioned and “rescue stations”, the EFDdb represents the process more detailedly, mentioning the activity of “squeezing out water” as well as “pipelines of steam pumps” and “harbour cranes”. As with the previous examples, the used UDC indices also do not match, EFDdb’s index representing the accident as an event in the sphere of engineering, while BIBIS locates it rather in the domain of communications.

5 Conclusion

As cultural resources of the past become digitised and stored in databases they are also getting “metadated”—framed by metadata—to use a verb borrowed from Wolfgang Ernst.²⁹ Following Jeffrey Pomerantz, we suggest that metadata schemas could be understood as “very simple languages” doing mediation work. Yet, in our article we have also tried to

²⁹ Wolfgang Ernst, *Digital Memory and the Archive*, University of Minnesota Press, 2013.

demonstrate that simplification itself is complex. We suggest that a metadata schema as a language functions in culture in semiotic terms as a form of metalanguage that models a domain of culture. As such, any model that is created is always a choice. It may emphasize specific features and leave others out. Furthermore, we have demonstrated that metadata systems could be understood as modelling multiple domains including the particularities of the object texts, the histories/cultures these object texts represent, our contemporary perceptions of these histories and, importantly the potential uses of the object texts in the present and in the imaginable future. That is, the modelling work a descriptive metadata schema is doing is effectively dialogic, interrelating and reconciling several different models.

Based on these theoretical assumptions, our empirical project ventured to comparatively analyse the modelling work of two databases—the Estonian Film Database (EFDb) and the Analytic Bibliography of Estonian Journalism (BIBIS)—in terms of how these model their respective modally different object corpuses (newsreels and articles of the newspaper *Postimees*) for the period from 1935 to 1939. Our article suggests that there are significant differences in previous modelling work. While EFDb is institutionally a new database aimed at mediating film heritage to all interested Internet users, BIBIS is effectively a path-dependent online version of how its parent institution, the Archival Library of Estonian Literary Museum, manages the resources it is set to preserve. This means that BIBIS is using the Estonian version of the international UDC standard to annotate all the individual newspaper articles. In our article we have accentuated the ways in which this standard is applied in BIBIS—with its generally high levels of abstraction when it comes to the modelling process—and pinpointed that it may easily mislead in case of online queries that in most cases presume higher precision in modelling.

In comparison we saw how the EFDb metadata schema tries to model in complex ways its audiovisual object corpus and the historical realities the films in the corpus show, a model that is also shaped by its expected uses. The eventual long list of keywords in EFDb indicate the challenge this complex set of goals presents to editors and annotators of the database. This could be deriving from the ambition to mark these different aspects in maximum detail—i.e. to create a metalanguage/model that describes all the inherent varieties of the object and its uses. Yet, the fact that 71 percent of its keywords were used only once suggests that the model is too “close” to its object—that it loses its generalizing function.

The example presents what is one of the core questions for metadata schemas: the appropriate balance between modelling precision, to deliver the most relevant results to a variety of “granular” online search queries, but also to generalize—to model structural connections between phenomena and domains. The dilemma is, as demonstrated by our analysis of the UDC-based modelling, that too general modelling process might become misleading. Furthermore, our study demonstrate that even high-level generalization as a simplification is a choice. The selections of UDC categories to model specific historical events by different databases were different. This could often (but not always) be due to medium-specific ways of modelling historical events—the political economies of their governance or the modal affordances of newsreels or written news articles make them show/tell different aspects of these events. But our examples also make salient the mediation processes by metadata schemas and their potential effects on the uses of these content databases, the reception of historical texts and, through this, broader processes of cultural memory formation in a culture. That is, at the time when many governments are investing large sums in digitisation of audiovisual heritage, when relevant databases and online services are created and metadata schemas get codified, it needs to be realised that these processes are in many ways convoluted. The time is hence ripe to start studying these modelling processes critically and in detail.

The writing of this article was supported by Estonian Research Council, grant PUT1176.

Biographies

Maarja Ojamaa is a research fellow at Tallinn University's (TLU) Baltic Film, Media, Arts and Communication School (BFM), Estonia. She is also a member of TLU Centre of Excellence in Media Innovation and Digital Culture (MEDIT).

She is additionally affiliated with the Department of Semiotics at the University of Tartu, where she defended her PhD thesis on the transmedial aspect of cultural autocommunication. Maarja's primary research interests lie in the transmedial mechanisms of cultural memory and in ways of diversifying the reusage of cultural heritage.

Indrek Ibrus is Professor of Media Innovation at Tallinn University's (TLU) Baltic Film, Media, Arts and Communication School (BFM), Estonia. He is also the Head of TLU Centre of Excellence in Media Innovation and Digital Culture (MEDIT). Indrek researches media innovation and evolution and combines cultural and economic approaches for this purpose. His main ongoing research project focuses on the role of metadata in contemporary cultural change.