



REPUBLIC OF ESTONIA  
MINISTRY OF EDUCATION  
AND RESEARCH



Horizon 2020  
European Union Funding  
for Research & Innovation

January 18, 2023

Authors: Mila Oiva, Hadi Nowandish, Helena Hanna Juht, Mariliis Niinemägi, Vejune Zemaityte, Ksenia Mukhina, Abida Bibi and Maximilian Schich

# Kinokroonika Data Description

The Kinokroonika dataset is composed of Estonian newsreel data digitized and stored in the Estonian Film Archive of the Estonian National Archives (EFA - <https://www.ra.ee/en/film-photo-audio/film/>) and the Estonian Film Database (EFDB - <https://www.efis.ee/>) that was enriched by the *Kinokroonika Exploration Project. Computational Data Enrichment and Distant Viewing of Estonian Newsreels (1922-1997)* project. The Kinokroonika project is funded by the *Estonian Language and Culture in the Digital Age (EKKD)*, a national program of the Ministry of Education and Research of the Republic of Estonia for April 2022-February 2023 (funding no EKKD77).

The dataset covers the main newsreels produced by *Eesti Kulturfilm* and *Tallinnfilm* studios in 1923-1997. The newsreel journals covered are *Eesti Kulturfilmi ringvaade*, *Ringvaade "Nõukogude Eesti"*, and *Eesti Kroonika*. The dataset contains over 2 100 newsreel issues from those years. The mean duration of the newsreels (of those we have length data for) is 2,23 min. This means that we have approximately 24 000 minutes of newsreel data.

For the basis of the metadata, we used EFDB data, and we received the digitized video files from EFA.

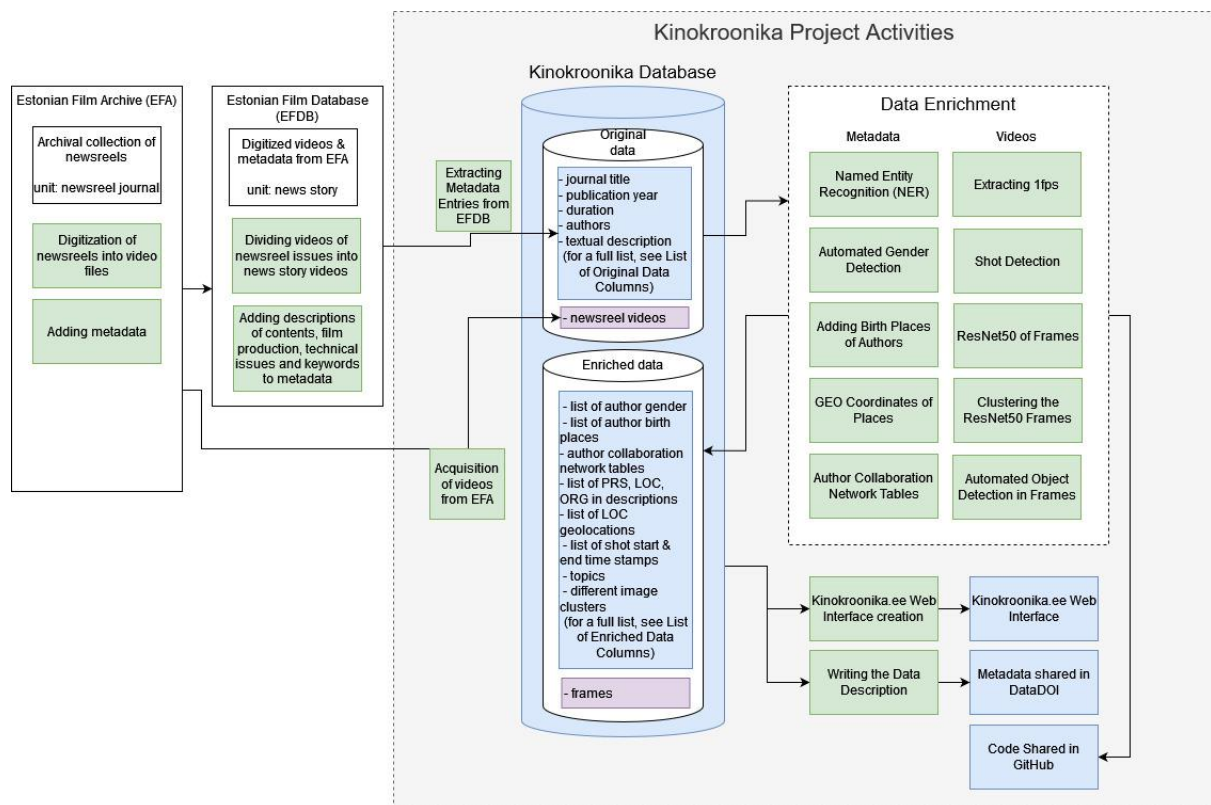


Image 1: Description of the flow of data and the activities related to the data processing in the Kinokroonika project.

[here about data sharing and licence agreements] We have licence agreement with the Estonian Film Institute, which is the copyright holder of the collection of newsreels produced by the Tallinnfilm film studios in 1940-1997. The licence grants our project a right to share the enriched data for research and educational purposes from August 2022 to August 2032. [What about the newsreels before that, what is the copyright situation with them?] [Add information concerning licence and data sharing agreements with the Estonian Film Archive and Estonian Film Database]

## Contents of the Data

The data contains news stories dating back from 1911 to 2015. There are news stories from several newsreel journals including the prominent *Eesti Kultuurifilmi ringvaade* [that was most likely the same as *Eesti Kultuurifilmi kroonika* and *Eesti Kultuurifilmi filmikroonika*] covering years 1923-1940, *Nõukogude Eesti (Soviet Estonia)* covering years 1940-1941 and 1944-1988, and *Eesti Filmikroonika* [that is most likely the same as *Eesti Kroonika*] covering years 1989-1998. In addition there are some less seldom titles, such as *Estonia-Filmi kroonika* (1913-1931), *Eduard Olari "Eesti Kroonika"* (1954-1961), *E.Prosti Filmikroonika* (1965-1980) and *Tallinnfilmi elu* (1967-1989). Under the coverage of "miscellaneous" we have listed journals with small titles, such as *Eesti Kultuurifilmi noorsoopalade montaaž* (1936 and 1938) and *Filmikroonika "Kuldne jõgi"* (1987).

Discrepancies, typos, and using different forms on newsreel journal titles causes that some titles run under different names in the data. Some metadata might be added to wrong places, leading to a situation where information cannot be identified although it is listed. The title "Filmikroonika"-- meaning literally "film chronicle" – is most likely just a generic title added for newsreel-like materials when adding the metadata. The title of the journal series is unknown (or placed somewhere else in the metadata) for altogether 2 872 news stories, which means 29% of the whole data. There are approximately 260 news stories without a year of publication.

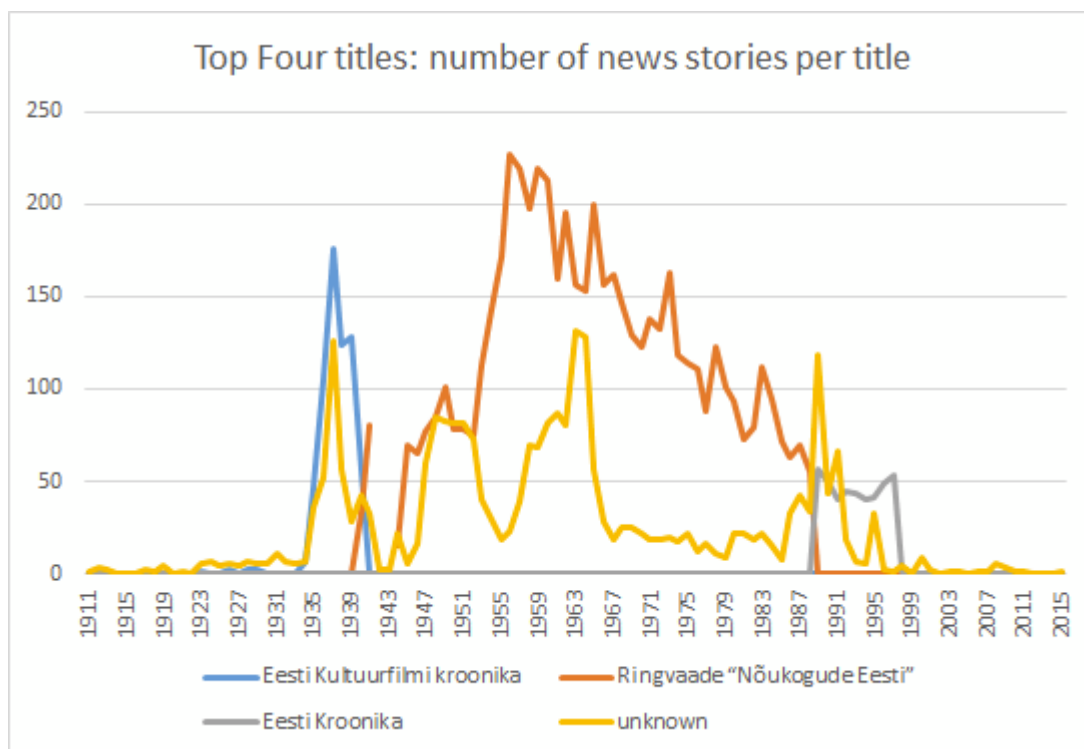
Journal name	total number	years
Estonia-Filmi kroonika	22	1913, 1921, 1923-1929, 1931
Filmoteegi materjal	41	1923-1928, 1935
Eesti Kulturfilmi ringvaade	612	1923, 1926, 1928-1929, 1934-1940
Eesti Kulturfilmi kroonika	25	1926, 1935-1938, 1940
Eesti Kulturfilmi filmikroonika	12	1935, 1938-1939
Die Deutsche Wochenschau	8	1940, 1944-1945
Eesti kroonika ja kirevat välisilmast	2	1937
Ringvaade "Nõukogude Eesti"	5642	1940-1941, 1944-1988
Eduard Olari "Eesti Kroonika"	31	1954, 1958-1961
E.Prosti Filmikroonika	49	1965-1970, 1972, 1980
Tallinnfilmi elu	35	1967-1968, 1970, 1972, 1975-1977, 1979-1987, 1989
Eesti Filmikroonika	10	1989, 1991, 1995
Eesti Kroonika	408	1989-1998
Filmikroonika	42	1919, 1924, 1939-1940, 1944-1949, 1973-1975, 1979, 1983, 1985, 1988
miscellaneous	91	randomly 1924-1993
unknown	2872	randomly 1911-2015
<b>all</b>	<b>9902</b>	<b>1911-2015</b>

Table 1: Number and publication years of news stories in the database listed based on the journal titles.

Visualizing the same information allows one to see the Estonian newsreel publication dynamics better. The heat map (Image 2) reveals that the news stories from the inter-war



unknown titles, we can see three distinct periods of newsreel production that correspond with the political changes in the country. Interestingly the peaks of unknown titles correspond with the rise of newsreel production in the mid 1930s, launch of the *Soviet Estonia* newsreel journal and launch of *Eesti Kroonika* in the late 1980s. It is possible to speculate that many of the unknown titles may have belonged to the three main titles listed here, and thus that the actual number of the main titles in the data can be larger than presented by these figures. Interestingly the number of unknown titles increased also in the late 1950s and early 1960s, when the number of Soviet Estonia news stories started to decrease. Is there a relation between them?



[how does the number of news stories look like if we look at the smaller titles? Are there interesting patterns?]

Evidently the number of newsreels from different years in the dataset does not necessarily correspond with the amount of newsreels actually produced. Some newsreels might have not ended in the archive. [Is there a way to check this - issue numbers and years!]

If we look at the all data stored in Film Archive of the National Archives of Estonia labelled as 'newsreels' (for details, see the sub-title "Extracting Metadata Entries from EFDB" below) we get to see an interesting timeline of different producing companies whose production has ended up in the archive. The visualization shows partially which companies have been

producing newsreels in Estonia at what time, but also productions of which companies have ended up for a variety of reasons in the archive. The reasons might be that the newsreels of the company have been circulating in Estonia, and ended up in the archive in that way, or some individual collectors might also have donated their collections to the archive. [check Närepea's article on the archival collections]

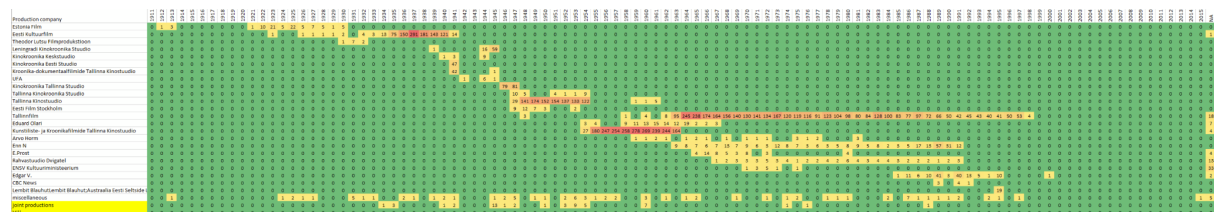


Image 5: Number of news stories by producing companies in 1911-2015 listed as newsreels in the collections of the Film Archive of the National Archives of Estonia.

Looking at the distribution of the collection already gives an interesting glimpse to Estonian newsreel production history. It shows parallel activities of *Estonia Film* and *Eesti Kulturfilm* in the 1920s. In the 1920s *Estonia Film* was more productive, but in the 1930s *Eesti Kulturfilm* won the competition and started to produce great numbers of newsreels, while *Estonia Film* vanished from the scene (Image 6). [Wikipedia and EFIS state that *Estonia Film* would have existed in 1919-1932, and *Eesti Kulturfilm* in 1931-1941,<sup>1</sup> which does not correspond with our data. Both also show earlier film production.] It also shows that a production company called *Theodor Lutsu Filmproduktioon* acted as a bridge in the transition in 1930-1932 taking care of newsreel production in Estonia. The owner of the company, Theodor Luts, was a known Estonian film-maker who had studied filmmaking in Paris and Berlin in the early 1920s. In 1932 he moved first to Finland, and later to Sweden and Brazil after World War II.<sup>2</sup>

Production company	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941	1942	1943	
Estonia Film	0	1	3	0	0	0	0	0	0	0	0	1	10	21	5	22	5	7	5	1	5	0	0	0	0	0	0	0	0	0	0	0	0	
Eesti Kulturfilm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	3	13	75	150	291	181	143	121	14	0	0	
Theodor Lutsu Filmproduktioon	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	7	2	0	0	0	0	0	0	0	0	0	0	0
Leningradi Kinokroonika Stuudio	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Kinokroonika Keskstuudio	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	0
Kinokroonika Eesti Stuudio	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	47	0	0
Kroonika-dokumentaalfilmide Tallinna Kinostuudio	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	42	0	0
UFA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Kinokroonika Tallinna Stuudio	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

<sup>1</sup> Wikipedia pages and EFDB on Estonia Film and Eesti Kulturfilm: <https://et.wikipedia.org/wiki/Estonia-Film>; [https://et.wikipedia.org/wiki/Eesti\\_Kulturfilm](https://et.wikipedia.org/wiki/Eesti_Kulturfilm); <https://www.efis.ee/et/filmitegijad/asutus/id/56/>; and <https://www.efis.ee/et/filmitegijad/asutus/id/46> (all accessed October 20, 2022).

<sup>2</sup> Eesti Entsüklopeedia on Theodor Luts [http://entsyklopeedia.ee/artikkel/luts\\_theodor](http://entsyklopeedia.ee/artikkel/luts_theodor) (accessed October 19, 2022); Estonian Film Database on Theodor Luts <https://www.efis.ee/et/varamu/filmitegijate-leksikon/le-li/theodor-luts> (accessed October 19, 2022).

Image 6: A closeup of the number of news stories by production companies in 1911-1943 in the collections of the Film Archive of the National Archives of Estonia.

When we focus on the companies producing newsreels in the years of World War II in Estonia we can see fluctuation and instability of the war times. We can see the Moscow based *Kinokroonika Keskstuudio* and Leningrad based *Leningrad Film Studio*, production of which was distributed during the first and at the beginning of the second Soviet occupation, and the German *UFA* distributing its production in 1942-1945. In addition there were two Estonian film studios – *Kinokroonika Eesti Stuudio* [this was the new name of the nationalized Eesti Kulturfilm<sup>3</sup>] and *Kroonika-dokumentaalfilmide Tallinna Kinostuudio* – [both of which were predecessors of Tallinnfilm<sup>4</sup>] that were producing newsreels during World War II.

Production company	1935	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945	1946	1947	1948
Estonia Film	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Eesti Kulturfilm	75	150	291	181	143	121	14	0	0	0	0	0	0	0
Theodor Lutsu Filmproduktioon	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Leningradi Kinokroonika Stuudio	0	0	0	0	1	0	0	0	0	16	59	0	0	0
Kinokroonika Keskstuudio	0	0	0	0	0	1	3	0	0	9	0	0	0	0
Kinokroonika Eesti Stuudio	0	0	0	0	0	0	47	0	0	0	0	0	0	0
Kroonika-dokumentaalfilmide Tallinna Kinostuudio	0	0	0	0	0	0	42	0	0	0	1	0	0	0
UFA	0	0	0	0	0	0	0	1	0	6	1	0	0	0
Kinokroonika Tallinna Stuudio	0	0	0	0	0	0	0	0	0	0	0	79	81	0
Tallinna Kinokroonika Stuudio	0	0	0	0	0	0	0	0	0	0	0	0	10	5
Tallinna Kinostuudio	0	0	0	0	0	0	0	0	0	0	0	0	29	141
Eesti Film Stockholm	0	0	0	0	0	0	0	0	0	0	0	0	9	12
Tallinnfilm	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Eduard Olari	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kunstiliste- ja Kroonikafilmi Tallinna Kinostuudio	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Arvo Horm	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Enn N	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E.Prost	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Rahvastuudio Dvigatel	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSV Kultuuriministeerium	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Edgar V.	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CBC News	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lembit BlauhutLembit Blauhut;Austraalia Eesti Seltside Li	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Image 7: A closeup of the number of news stories produced in 1935-1948 with production companies active in the World War II era in Estonia.

A visualisation of the newsreel companies present in the database after World War II shows how during the post-war years the name of the main newsreel producing company was changed. From 1945 to 1954 it was called the *Tallinn Studio of Newsreels (Kinokroonika*

<sup>3</sup> The Estonian Film Database on Kinokroonika Eesti Stuudio <https://www.efis.ee/et/filmitegijad/asutus/id/960/> (accessed October 21, 2022). Our data shows that they produced 47 news stories of Ringvaade "Nõukogude Eesti" in 1941.

<sup>4</sup> Wikipedia on Tallinnfilm <https://et.wikipedia.org/wiki/Tallinnfilm> (accessed in October 21, 2022).





Image 9: A closeup of the number of news stories produced in 1946-2000 studios producing smaller series of (documentary) films highlighted.

Exclusions from the dataset: These film producers are mostly individual or amateur filmmakers and we have therefore excluded these films. The collection also contains 51 joint productions of two film studios, which were predominantly joint productions of an Estonian film studio and a studio from another Soviet republic during the 1940s to 1960s. In addition, the dataset contained 86 films produced by miscellaneous companies that did not compose series of (documentary) film production in the Estonian data. We have excluded these from our dataset.

## Construction of the Dataset

### Extracting Metadata Entries from EFDB

We chose to use the EFDB metadata as a basis, because EFA and EFDB metadata overlap to a great extent, but EFDB data has some details more than EFA metadata. [The basis of the metadata was added in the EFA during the digitization process on the basis of older archival records.[?]] A difference between these metadata sets is that in EFA the basic unit is one newsreel issue (an approximately 10 min long news film that contains 5-7 news stories) while in EFDB the basic unit is a news story. EFDB has also added more detailed descriptions of contents, film production, technical issues and keywords related to picture and text. The leader of the EFDB metadata enrichment at the Estonian Film Database Non-Governmental Organization was Visiting Professor of Television Culture at Tallinn University Hagi Šein.

The EFDB database is a [relational database in MySQL format that allows exporting data in xml form. Altogether, it consists of x number of separate but interlinked tables] To extract newsreels from the EFIS database, we used the following film\_type\_id categories: 110, 59, 21, 96 and 97 [what are the differences between these film categories?].

	Film type id	language	title	MATCHES in efis.efis_film_type_re l_veebist table	MATCHES in efis.efis_film_type_re l table
213	110	en	Newsreel	0	1,043
67	59	en	Newsreels	2,909	8,742
64	21	en	Newsreels 1912-1940 I half	380	1,125
155	96	en	Newsreels 1940 II half-1991	394	7,312
157	97	en	Newsreels beyond 1992	203	286
			Total = 22,394	3,886	18,508
				Unique total*	9,902

\*Some films are tagged with multiple newsreel tags, such as 21 & 59. Others appear in both tables (veebist and not). Overall, there are 9,902 unique newsreels with metadata in the database.

Film lists were extracted from both tables based on film type. Those lists were merged with duplicate film IDs removed to arrive at a list of unique IDs.

Additional data for that list of films was then extracted.

The EFDB database consisted of all together 38 separate tables containing information from film contents, music, authors, to technical details and distribution, and archiving. Because the purpose of the database is to facilitate research on depictions of world views in newsreels, we selected only part of the contents and collapsed the data into a long csv table. (For a more detailed list of all the EFDB data entries, and the ones selected to Kinokroonika dataset, please see Appendix 1 - List of all EFDB tables and columns and the ones selected to Kinokroonika.) We extracted the necessary EFDB data entries as xml files and collapsed the data into a long csv table using the following steps:

1. We have data in xml files.
2. We converted in to csv files
3. We read the csv files 3 files were not readable (anonymous error)
4. We converted that 34 files into xls files to make them readable.
5. After that we started extracting useful columns.

Few columns were selected from each data file that are mentioned below. Few columns are added as it is from efis data files few were just renamed as per need.

List of column titles / metadata titles in Master.csv extracted from the EFIS database:

- [1] "film\_id"
- [2] "film\_seq\_id"

- [3] "production\_year"
- [4] "film\_presentation\_lang"
- [5] "film\_voiceover\_lang"
- [6] "film\_subtitles\_lang"

added contents to these new columns from column "language\_code" by extracting the ending signifying the language "ET", "EN", "RU" etc (for example "FILM\_LANGUAGE\_ET" -> "ET")

- [7] "title"
- [8] "parallel\_title"

- added contents from column "title" of this table to this new column corresponding to the film\_id

- [9] "film\_physical\_duration"
- [10] "film\_physical\_metrage"
- [11] "film\_physical\_notes"

- added contents from column "classifier\_code2" -> "FILM\_PHYSICAL\_TYPE\_FILM\_FILM\_DURATION", "FILM\_PHYSICAL\_TYPE\_FILM\_FILM\_METRAGE", "FILM\_PHYSICAL\_TYPE\_FILM\_FILM\_NOTES" -> content from column "amount"

- [12] "country\_id"

(add contents from column "country\_id" -> we need to replace the country\_id numbers with the names of countries)

- [13] "production\_companies"
- [14] "coproducers"
- [15] "sound\_editors"

(added contents from column "company\_name" of this table to the new column corresponding the film\_id)

- [16] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_film\_music\_id"
- [17] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_film\_music\_id"
- [18] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_film\_music\_title"
- [19] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_film\_music\_title"
- [20] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_film\_music\_seq"
- [21] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_film\_music\_seq"
- [22] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_company\_id"
- [23] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_company\_id"
- [24] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_people\_id"
- [25] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_people\_id"
- [26] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_film\_music\_maker\_seq"
- [27] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_film\_music\_maker\_seq"
- [28] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_full\_name"
- [29] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_full\_name"
- [30] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_birth\_death"
- [31] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_birth\_death"
- [32] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_people\_num\_related\_films"
- [33] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_people\_num\_related\_films"

[34] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_people\_notes"  
[35] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_people\_notes"  
[36] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_company\_name"  
[37] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_company\_name"  
[38] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_country\_id"  
[39] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_country\_id"  
[40] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_company\_type"  
[41] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_company\_type"  
[42] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_company\_num\_related\_films"  
[43] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_company\_num\_related\_films"  
[44] "film\_variation"  
[45] "film\_version"

(added contents from column "title" corresponding to  
"FILM\_EDITION\_TYPE\_VARIATION", "FILM\_EDITION\_TYPE\_VERSION" in column  
"classificator\_code")

[46] "screening\_cinema\_loc"  
[47] "screening\_dvd\_loc"  
[48] "screening\_p\_loc"  
[49] "screening\_premiere\_loc"  
[50] "screening\_other\_loc"  
[51] "screening\_festival\_loc"  
[52] "screening\_television\_loc"  
[53] "screening\_cinema\_date"  
[54] "screening\_dvd\_date"  
[55] "screening\_p\_date"  
[56] "screening\_premiere\_date"  
[57] "screening\_other\_date"  
[58] "screening\_festival\_date"  
[59] "screening\_television\_date"

(added contents to the new columns from columns "location" (for new columns  
ending "\_loc") and "event\_info" (for new columns ending with "\_date") corresponding  
to the film\_id.

)

[60] "relation\_film"  
[61] "relation\_series"  
[62] "people\_id\_CINEMATOGRAPHER"  
[63] "people\_id\_DIRECTOR"  
[64] "people\_id\_SCREENWRITER"  
[65] "people\_id\_CONTENT\_EDITOR"  
[66] "people\_id\_SOUND\_RECORDIST\_PRODUCTION\_SOUND\_MIXER"  
[67] "people\_id\_SOUND\_DESIGNER"  
[68] "people\_id\_COMPOSER"  
[69] "people\_id\_SOUND\_DESIGNER2"  
[70] "people\_id\_TEXT\_AUTHOR"  
[71] "people\_id\_EDITOR"



[118] "full\_name\_ELDER\_ADMINISTRATOR"  
[119] "full\_name\_LIGHTER"  
[120] "full\_name\_TRICARTIST"  
[121] "full\_name\_CAR\_DRIVER"  
[122] "full\_name\_2\_ASSITING\_CAMERA"  
[123] "full\_name\_PRODUCTION\_SOUND\_MIXER"  
[124] "full\_name\_TRIC\_CAMERA"  
[125] "full\_name\_WRITER\_OF\_DIALOGUE"  
[126] "birth\_death\_CINEMATOGRAPHER"  
[127] "birth\_death\_DIRECTOR"  
[128] "birth\_death\_SCREENWRITER"  
[129] "birth\_death\_CONTENT\_EDITOR"  
[130] "birth\_death\_SOUND\_RECORDIST\_PRODUCTION\_SOUND\_MIXER"  
[131] "birth\_death\_SOUND\_DESIGNER"  
[132] "birth\_death\_COMPOSER"  
[133] "birth\_death\_CONDUCTOR"  
[134] "birth\_death\_SOUND\_DESIGNER2"  
[135] "birth\_death\_TEXT\_AUTHOR"  
[136] "birth\_death\_EDITOR"  
[137] "birth\_death\_MANAGER"  
[138] "birth\_death\_ANNOUNCER"  
[139] "birth\_death\_1\_ASSISTANT\_CAMERA"  
[140] "birth\_death\_MUSIC\_DESIGNER"  
[141] "birth\_death\_CONSULTANT"  
[142] "birth\_death\_STAGE\_DIRECTOR"  
[143] "birth\_death\_ADMINISTRATOR"  
[144] "birth\_death\_ASSOCIATE\_DIREKTOR\_AD"  
[145] "birth\_death\_MULTIOPERATOR"  
[146] "birth\_death\_NARRATOR"  
[147] "birth\_death\_TRICK\_CAMERA"  
[148] "birth\_death\_TRANSLATOR"  
[149] "birth\_death\_NEGATIVE\_CUTTER"  
[150] "birth\_death\_ELDER\_ADMINISTRATOR"  
[151] "birth\_death\_LIGHTER"  
[152] "birth\_death\_AUDIO\_TECHNICIAN"  
[153] "birth\_death\_PRODUCTION\_SOUND\_MIXER"  
[154] "birth\_death\_SOUND\_RECORDING\_\_ASSITANT"  
[155] "birth\_death\_LIGHTING\_DIRECTOR\_LD\_GAFFER"  
[156] "birth\_death\_PRODUCER"  
[157] "birth\_death\_WRITER"  
[158] "birth\_death\_ANIMATOR"  
[159] "birth\_death\_AUTHOR"  
[160] "birth\_death\_RE\_RECORDING\_MIXER\_DUBBING\_MIXER"  
[161] "birth\_death\_WRITER\_OF\_DIALOGUE"  
[162] "num\_related\_films\_CINEMATOGRAPHER"  
[163] "num\_related\_films\_DIRECTOR"



- [210] "notes\_NARRATOR"
- [211] "notes\_TRANSLATOR"
- [212] "notes\_ELDER\_ADMINISTRATOR"
- [213] "notes\_LIGHTER"
- [214] "notes\_PRODUCTION\_SOUND\_MIXER"
- [215] "keyword\_values\_et"
- [216] "keyword\_values\_en"
- [217] "keyword\_group\_et"
- [218] "keyword\_group\_en"
- added contents from the EFIS table
  
- [219] "language"
- [220] "title\_proper"
- [221] "title\_alternative"
- [222] "title\_avant"
- [223] "title\_working"
- [224] "title\_serial"
- [225] "title\_additional\_info"
- [226] "issue\_no"
- column "title\_alternative", from EFIS data ,added the number following "nr" in this column – (for example "Ringvaade vTõhNГµkogude EestivTõk nr 6 (715), veebruar 1963" -> add "6" or "Ringvaade vTõhNГµkogude EestivTõk nr 1 (1277), 1/6" -> added "1")

## List of Original Data Columns (Base Master.csv)

- [1] "film\_id"
- [2] "film\_seq\_id"
- [3] "archive\_fond\_nr"
- [4] "archive\_archival\_document\_nr"
- [5] "production\_year"
- [6] "film\_presentation\_lang"
- [7] "film\_voiceover\_lang"
- [8] "film\_subtitles\_lang"
- [9] "title"
- [10] "parallel\_title"
- [11] "film\_physical\_duration"
- [12] "film\_physical\_metrage"
- [13] "film\_physical\_notes"
- [14] "country\_id"
- [15] "country\_name\_en\_title"
- [16] "country\_name\_est\_title"
- [17] "production\_companies"
- [18] "coproducers"



[19] "sound\_editors"  
[20] "language"  
[21] "title\_proper"  
[22] "title\_alternative"  
[23] "title\_avant"  
[24] "title\_working"  
[25] "title\_serial"  
[26] "title\_additional\_info"  
[27] "issue\_no"  
[28] "relation\_film"  
[29] "relation\_series"  
[30] "keyword\_values\_et"  
[31] "keyword\_values\_en"  
[32] "keyword\_group\_et"  
[33] "keyword\_group\_en"  
[34] "film\_variation"  
[35] "film\_version"  
[36] "screening\_cinema\_loc"  
[37] "screening\_dvd\_loc"  
[38] "screening\_p\_loc"  
[39] "screening\_premiere\_loc"  
[40] "screening\_other\_loc"  
[41] "screening\_festival\_loc"  
[42] "screening\_television\_loc"  
[43] "screening\_cinema\_date"  
[44] "screening\_dvd\_date"  
[45] "screening\_p\_date"  
[46] "screening\_premiere\_date"  
[47] "screening\_other\_date"  
[48] "screening\_festival\_date"  
[49] "screening\_television\_date"  
[50] "people\_id\_CINEMATOGRAPHER"  
[51] "people\_id\_DIRECTOR"  
[52] "people\_id\_SCREENWRITER"  
[53] "people\_id\_CONTENT\_EDITOR"  
[54] "people\_id\_SOUND\_RECORDIST\_PRODUCTION\_SOUND\_MIXER"  
[55] "people\_id\_SOUND\_DESIGNER"  
[56] "people\_id\_COMPOSER"  
[57] "people\_id\_SOUND\_DESIGNER2"  
[58] "people\_id\_TEXT\_AUTHOR"  
[59] "people\_id\_EDITOR"  
[60] "people\_id\_....."  
[61] "people\_id\_MANAGER"  
[62] "people\_id\_ANNOUNCER"  
[63] "people\_id\_1\_ASSISTANT\_CAMERA"  
[64] "people\_id\_MUSIC\_DESIGNER"  
[65] "people\_id\_CONSULTANT"

[66] "people\_id\_STAGE\_DIRECTOR"  
[67] "people\_id\_ADMINISTRATOR"  
[68] "people\_id\_ASSOCIATE\_DIREKTOR\_AD"  
[69] "people\_id\_MULTIOPERATOR"  
[70] "people\_id\_NARRATOR"  
[71] "people\_id\_SOUND\_EDITOR"  
[72] "people\_id\_TRICK\_CAMERA"  
[73] "people\_id\_TRANSLATOR"  
[74] "people\_id\_NEGATIVE\_CUTTER"  
[75] "people\_id\_ELDER\_ADMINISTRATOR"  
[76] "people\_id\_LIGHTER"  
[77] "people\_id\_TRICARTIST"  
[78] "people\_id\_CAR\_DRIVER"  
[79] "people\_id\_2\_ASSITING\_CAMERA"  
[80] "people\_id\_PRODUCTION\_SOUND\_MIXER"  
[81] "people\_id\_REPORTER"  
[82] "full\_name\_CINEMATOGRAPHER"  
[83] "full\_name\_DIRECTOR"  
[84] "full\_name\_SCREENWRITER"  
[85] "full\_name\_CONTENT\_EDITOR"  
[86] "full\_name\_SOUND\_RECORDIST\_PRODUCTION\_SOUND\_MIXER"  
[87] "full\_name\_SOUND\_DESIGNER"  
[88] "full\_name\_COMPOSER"  
[89] "full\_name\_SOUND\_DESIGNER2"  
[90] "full\_name\_TEXT\_AUTHOR"  
[91] "full\_name\_EDITOR"  
[92] "full\_name\_....."  
[93] "full\_name\_MANAGER"  
[94] "full\_name\_ANNOUNCER"  
[95] "full\_name\_1\_ASSISTANT\_CAMERA"  
[96] "full\_name\_MUSIC\_DESIGNER"  
[97] "full\_name\_CONSULTANT"  
[98] "full\_name\_STAGE\_DIRECTOR"  
[99] "full\_name\_ADMINISTRATOR"  
[100] "full\_name\_ASSOCIATE\_DIREKTOR\_AD"  
[101] "full\_name\_MULTIOPERATOR"  
[102] "full\_name\_NARRATOR"  
[103] "full\_name\_SOUND\_EDITOR"  
[104] "full\_name\_TRANSLATOR"  
[105] "full\_name\_NEGATIVE\_CUTTER"  
[106] "full\_name\_ELDER\_ADMINISTRATOR"  
[107] "full\_name\_LIGHTER"  
[108] "full\_name\_TRICARTIST"  
[109] "full\_name\_CAR\_DRIVER"  
[110] "full\_name\_2\_ASSITING\_CAMERA"  
[111] "full\_name\_PRODUCTION\_SOUND\_MIXER"  
[112] "full\_name\_TRIC\_CAMERA"

[113] "full\_name\_WRITER\_OF\_DIALOGUE"  
[114] "birth\_death\_CINEMATOGRAPHER"  
[115] "birth\_death\_DIRECTOR"  
[116] "birth\_death\_SCREENWRITER"  
[117] "birth\_death\_CONTENT\_EDITOR"  
[118] "birth\_death\_SOUND\_RECORDIST\_PRODUCTION\_SOUND\_MIXER"  
[119] "birth\_death\_SOUND\_DESIGNER"  
[120] "birth\_death\_COMPOSER"  
[121] "birth\_death\_CONDUCTOR"  
[122] "birth\_death\_SOUND\_DESIGNER2"  
[123] "birth\_death\_TEXT\_AUTHOR"  
[124] "birth\_death\_EDITOR"  
[125] "birth\_death\_MANAGER"  
[126] "birth\_death\_ANNOUNCER"  
[127] "birth\_death\_1\_ASSISTANT\_CAMERA"  
[128] "birth\_death\_MUSIC\_DESIGNER"  
[129] "birth\_death\_CONSULTANT"  
[130] "birth\_death\_STAGE\_DIRECTOR"  
[131] "birth\_death\_ADMINISTRATOR"  
[132] "birth\_death\_ASSOCIATE\_DIREKTOR\_AD"  
[133] "birth\_death\_MULTIOPERATOR"  
[134] "birth\_death\_NARRATOR"  
[135] "birth\_death\_TRICK\_CAMERA"  
[136] "birth\_death\_TRANSLATOR"  
[137] "birth\_death\_NEGATIVE\_CUTTER"  
[138] "birth\_death\_ELDER\_ADMINISTRATOR"  
[139] "birth\_death\_LIGHTER"  
[140] "birth\_death\_AUDIO\_TECHNICIAN"  
[141] "birth\_death\_PRODUCTION\_SOUND\_MIXER"  
[142] "birth\_death\_SOUND\_RECORDING\_ASSITANT"  
[143] "birth\_death\_LIGHTING\_DIRECTOR\_LD\_GAFFER"  
[144] "birth\_death\_PRODUCER"  
[145] "birth\_death\_WRITER"  
[146] "birth\_death\_ANIMATOR"  
[147] "birth\_death\_AUTHOR"  
[148] "birth\_death\_RE\_RECORDING\_MIXER\_DUBBING\_MIXER"  
[149] "birth\_death\_WRITER\_OF\_DIALOGUE"  
[150] "num\_related\_films\_CINEMATOGRAPHER"  
[151] "num\_related\_films\_DIRECTOR"  
[152] "num\_related\_films\_SCREENWRITER"  
[153] "num\_related\_films\_CONTENT\_EDITOR"  
[154] "num\_related\_films\_SOUND\_RECORDIST\_PRODUCTION\_SOUND\_MIXER"  
[155] "num\_related\_films\_SOUND\_DESIGNER"  
[156] "num\_related\_films\_COMPOSER"  
[157] "num\_related\_films\_SOUND\_DESIGNER2"  
[158] "num\_related\_films\_TEXT\_AUTHOR"  
[159] "num\_related\_films\_EDITOR"

[160] "num\_related\_films\_....."  
[161] "num\_related\_films\_MANAGER"  
[162] "num\_related\_films\_ANNOUNCER"  
[163] "num\_related\_films\_1\_ASSISTANT\_CAMERA"  
[164] "num\_related\_films\_MUSIC\_DESIGNER"  
[165] "num\_related\_films\_CONSULTANT"  
[166] "num\_related\_films\_STAGE\_DIRECTOR"  
[167] "num\_related\_films\_ADMINISTRATOR"  
[168] "num\_related\_films\_ASSOCIATE\_DIREKTOR\_AD"  
[169] "num\_related\_films\_MULTIOPERATOR"  
[170] "num\_related\_films\_NARRATOR"  
[171] "num\_related\_films\_SOUND\_EDITOR"  
[172] "num\_related\_films\_TRANSLATOR"  
[173] "num\_related\_films\_NEGATIVE\_CUTTER"  
[174] "num\_related\_films\_ELDER\_ADMINISTRATOR"  
[175] "num\_related\_films\_LIGHTER"  
[176] "num\_related\_films\_TRICARTIST"  
[177] "num\_related\_films\_CAR\_DRIVER"  
[178] "num\_related\_films\_2\_ASSITING\_CAMERA"  
[179] "num\_related\_films\_PRODUCTION\_SOUND\_MIXER"  
[180] "num\_related\_films\_TRIC\_CAMERA"  
[181] "notes\_CINEMATOGRAPHER"  
[182] "notes\_DIRECTOR"  
[183] "notes\_SCREENWRITER"  
[184] "notes\_CONTENT\_EDITOR"  
[185] "notes\_SOUND\_RECORDIST\_PRODUCTION\_SOUND\_MIXER"  
[186] "notes\_SOUND\_DESIGNER"  
[187] "notes\_COMPOSER"  
[188] "notes\_SOUND\_DESIGNER2"  
[189] "notes\_TEXT\_AUTHOR"  
[190] "notes\_EDITOR"  
[191] "notes\_MANAGER"  
[192] "notes\_ANNOUNCER"  
[193] "notes\_1\_ASSISTANT\_CAMERA"  
[194] "notes\_MUSIC\_DESIGNER"  
[195] "notes\_ADMINISTRATOR"  
[196] "notes\_ASSOCIATE\_DIREKTOR\_AD"  
[197] "notes\_MULTIOPERATOR"  
[198] "notes\_NARRATOR"  
[199] "notes\_TRANSLATOR"  
[200] "notes\_ELDER\_ADMINISTRATOR"  
[201] "notes\_LIGHTER"  
[202] "notes\_PRODUCTION\_SOUND\_MIXER"  
[203] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_film\_music\_id"  
[204] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_film\_music\_id"  
[205] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_film\_music\_title"  
[206] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_film\_music\_title"

[207] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_film\_music\_seq"  
[208] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_film\_music\_seq"  
[209] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_company\_id"  
[210] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_company\_id"  
[211] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_people\_id"  
[212] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_people\_id"  
[213] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_film\_music\_maker\_seq"  
[214] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_film\_music\_maker\_seq"  
[215] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_full\_name"  
[216] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_full\_name"  
[217] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_birth\_death"  
[218] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_birth\_death"  
[219] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_people\_num\_related\_films"  
[220] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_people\_num\_related\_films"  
[221] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_people\_notes"  
[222] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_people\_notes"  
[223] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_company\_name"  
[224] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_company\_name"  
[225] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_country\_id"  
[226] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_country\_id"  
[227] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_company\_type"  
[228] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_company\_type"  
[229] "FILM\_MUSIC\_MAKER\_TYPE\_COMPOSER\_company\_num\_related\_films"  
[230] "FILM\_MUSIC\_MAKER\_TYPE\_PERFORMER\_company\_num\_related\_films"

The base Master.csv created in this way contains errors resulting from human coders adding the metadata in incorrect places.

## Acquisition of videos from EFA

[add description on how we acquired the videos, in what format]

## Data Enrichment

### Metadata Enrichment

[we still need to add detailed description of how the enrichment was done]

## For Named Entity Recognition [Oct 10th]

- Please check once again that we don't have in the EFIS tables any additional columns that contain additional textual descriptions of the films that we have not added to the Master.csv -> if yes, please contact Mila; if no, proceed to the next step
- Please run Named Entity Recognition for the following columns:
  - "title\_proper" -> NER in Estonian language
  - "keyword\_values\_et" -> NER in Estonian language
  - "keyword\_values\_en" -> NER in English language
  - "parallel\_title" -> NER in Russian? (I assume the column contents are in Russian, because the contents are currently in a non-understandable form, such as "PõP° PēPsP»C...PsP·PSC<C... PīPsP»CļC...")
  - "title" -> NER in Estonian language
  - Create new columns for each NER results and add the NER results into columns "title\_proper\_NER", "keyword\_values\_et\_NER", "keyword\_values\_en\_NER", "parallel\_title\_NER" and "title\_NER"

## Correcting the automated detection of place names

## Automated Gender Detection of authors and persons detected by NER [Oct 20th]

## Adding birthplaces of the authors

### Preparations by Hadi:

- The "unique\_authors" column comes from "authors\_list\_unique.csv".
- The "detected\_gender" column -> add results of automated gender detection
- Add empty column titles "correct\_name", "correct\_gender", "birthplace\_manual", "source\_manual".

### Adding information by Helena and Mariliis:

- "correct\_name": if the name in "unique\_authors" is correct -> leave empty; if the name in "unique\_authors" is not correct or is not complete -> add the correct or complete name. If the name in the "unique\_authors" looks normal and it's complete, you can rely on this information and you don't need to check that elsewhere.
- "correct\_gender": if the gender in "detected\_gender" is correct -> leave empty, if it is incorrect -> add correct gender. If you can see directly from the name the assumed gender, you don't need to check this information from elsewhere.

- “Birthplace\_manual”: please check the birthplace and add it here if possible. You can search for information for example:
  - By google (or yandex.ru can work sometimes)
  - Estonian Biographic database: <http://www2.kirmus.ee/biblioserver/isik/index.php?m=indexes&iid=20>
  - subject specific e-archives:
  - <https://www.ra.ee/dgs/explorer.php>
  -
- “source\_manual”: if you add contents in “birthplace\_manual”, add the source of information here (link, or name of the book).

In the manual processing of the author names, internet databases, internet searchers and manual library searchers were conducted. Internet searches and database searches proved to be fruitful whereas less biographical data could be acquired from books on newsreels and Estonian films. We noticed that it is not always easy to find the birthplaces of individuals, especially for filmmakers working with the earlier newsreels. Full contact info was not always available so an identification of first names for the persons was necessary in at least 83 cases. In some instances people working in the film industry would have the same name, this was more common for filmmakers with Russian names. A person might also have changed their name, either due to marriage or ‘Estonification’ of foreign surnames. In some cases, filmmakers used aliases for their work. There are approximately 20 persons with several names in the dataset.

It became apparent that some professions are better represented in the online resources than others. For example, drivers, lightning personnel and montage (that was mainly women’s work) are not often well represented in online resources and it is harder to find information on them, while sound engineers are better represented in the resources. Professions like text writers (experts on the field) are quite well represented in online searches but more commonly through their other endeavours. Earlier newscasters did have a long list of films they’re associated with but less is found of their personal data.

### Geo Coordinates of Places [Oct 28]

We added geocodes for the place names identified with the help of Named Entity Recognition. [which codes?] We counted one place name once for each news story. So, if, for example, Tallinn has been mentioned several times in one news story, we counted that only once. If several place names were mentioned in one news story, we counted all of them once. So the map visualizations show in how many news stories a certain place name has been mentioned, and not for how many times a place has been mentioned within one story. The logic behind is that a news story description can mention the location several times if the story is about an event in one city, but it is more important to count the number of news stories that mention a place.

The outcome of algorithms used for detecting geolocations, longitudes and latitudes was more or less correct. We had altogether 1411 identified locations (includes duplicates). However, there were some obvious errors in the results.

In these cases, we fixed the results manually. For that, we created a table of all the mentioned place names in order of higher to lowest frequency. Just to understand the different levels of place names, each place name was identified either as an exact place, district, village, borough, parish, city, river, lake, country or region.

Sometimes the algorithm did not recognize the Estonian place names if it's in another case than nominative. We corrected the case manually. Both English and Estonian correct versions of the place names were added to the table. In case the mentioned name had a newer version (such as Leningrad and Saint Petersburg), the new one was used and the old one was added to a separate column.

If a place name could signify multiple locations, then it was just marked as "multiple location" (about 7%). If the place name was not in fact a place name, then it was marked as "not a location" (about 20%). The latter were deleted from the table.

In addition, some results consisted of two place names, for example "Tallinn-Monte Carlo". In such cases we separated the two place names. If the frequency of "Tallinn-Monte Carlo" was 12, then both separated place names got the frequency of 12.

The final data table included ..... place names.

Creating two new spreadsheets with the places:

1. A spreadsheet containing only places that can be pointed with exact location (for the map in the web interface). This list includes the following place categories:
  - a. City
  - b. Island
  - c. District
  - d. Borough
  - e. County
  - f. Village
  - g. Place
  - h. Parish
2. A spreadsheet that contains all the data except for things that are not locations (data to be shared in DataDOI). This list includes in addition to the place categories listed in the first spreadsheet als
  - a. Country
  - b. River
  - c. Lake
  - d. Region
  - e. Mountain
  - f. Multiple locations



Author collaboration network tables [Nov 4th]

## Videos

Extracting image and numerical information from the videos

[Video] Extract 1 fps

Extracting one frame per second

[Video] Contrast identification

[Video] Shot detection

- Add column on shot lengths

[Video] Adding missing news story lengths?

[Video] ResNet50 of mid frames of each shot

[Video] Clustering the ResNet50 of mid frames of each shot

[Video] Add colour frames to different image clusters

[Video] Automated object detection in frames

## List of Enriched Data Columns

[add a complete list here]

- Named Entity Recognition (person, location, organisation) of descriptive columns
- Gender of authors (cinematographers and directors) and persons mentioned in the newsreels
- Birth places of the authors
- Geolocations (longitude and latitude) of placenames
- Author collaboration network data (gender, birthplace) - node and edge lists
- Extracted frames (24 fps)
- Shot detection
- 

## Appendix 1 - List of all EFDB tables and columns and the ones selected to Kinokroonika

This is a list of all the tables and columns of the Estonian Film Database (EFDB) containing information on newsreels. We have highlighted with **green** those columns that were selected to Kinokroonika dataset, and with **red letter colour** those ones that we were planning to add to Kinokroonika dataset, but that were empty or contained less than 1 000 values.

1. **Efis.efis\_film** - extracted as is, columns in **xxx** removed as empty or containing few values (under 1k)
  - a. id** - this is the unique film identifier
  - b. film\_seq\_id**
  - c. film\_version\_number**
  - d. production\_year**
  - e. is\_production\_year\_assumed**
  - f. profile\_image\_filename**

- g. copyright\_type
- h. copyright\_register\_info
- i. copyright\_note\_on\_movie
- j. archive\_standard\_number
- k. archive\_kriimsilm\_id
- l. archive\_movie\_archive\_url
- m. archive\_err\_url
- n. archive\_arrival\_date
- o. archive\_handover\_person\_name
- p. archive\_fond\_name
- q. archive\_fond\_nr
- r. archive\_archival\_document\_nr
- s. archive\_usage\_copy
- t. archive\_usage\_format
- u. archive\_usage\_start\_date
- v. archive\_usage\_end\_date
- w. archive\_access\_restriction
- x. archive\_additional\_material
- y. record\_info
- z. record\_sources
- aa. record\_creator\_opinion
- bb. efa\_editor\_opinion
- cc. efa\_editor\_status
- dd. is\_latest
- ee. is\_visible
- ff. x\_road\_synced
- gg. status\_by
- hh. status
- ii. status\_datetime
- jj. updated
- kk. updated\_by
- ll. created
- mm. Created\_by

## 2. Efis.efis\_film\_copyright

- a. id
- b. film\_id
- c. classifier\_code
- d. company\_id
- e. people\_id
- f. seq
- g. created
- h. created\_by
- i. updated
- j. updated\_by

3. Efis.efis\_film\_country - extracted as is, all data is full, needs to be connected to country via id

- a. id
- b. film\_id
- c. country\_id
- d. seq
- e. created
- f. created\_by
- g. updated
- h. updated\_by

4. efis.efis\_film\_distributor - extracted as is, empty columns removed, connected to company via id

- a. id
- b. film\_id
- c. language
- d. classifier\_code
- e. company\_id
- f. distribution\_info
- g. seq
- h. created
- i. created\_by
- j. updated
- k. Updated\_by

5. Efis.efis\_company - needs to be connected to country via id, empty columns removed

- a. id
- b. name
- c. country\_id
- d. type
- e. year\_founded
- f. ark\_code
- g. is\_visible
- h. num\_related\_films
- i. notes
- j. created
- k. created\_by
- l. updated
- m. updated\_by

6. Efis.efis\_film\_edition - extracted as is

- a. id
- b. film\_id
- c. lang
- d. classifier\_code
- e. title

- f. seq
- g. updated
- h. updated\_by
- i. created
- j. Created\_by

7. **Efis.efis\_film\_maker** - extracted with people information from efis\_people

- a. id
- b. film\_id
- c. classifier\_code
- d. people\_id
- e. seq
- f. created
- g. Created\_by

8. **Efis.efis\_people** - extracted for cast

- a. id
- b. first\_name
- c. last\_name
- d. date\_of\_birth
- e. date\_of\_death
- f. is\_visible
- g. num\_related\_films
- h. notes
- i. created
- j. created\_by
- k. updated
- l. Updated\_by

9. **Efis.efis\_film\_music** - connected to music maker and music making company

- a. id
- b. film\_id
- c. title
- d. seq
- e. created
- f. created\_by
- g. updated
- h. updated\_by

10. **Efis.efis\_film\_music\_maker**

- a. id
- b. film\_music\_id
- c. classifier\_code
- d. company\_id
- e. people\_id
- f. seq
- g. created
- h. created\_by

- i. updated
  - j. updated\_by
11. **Efis.efis\_people**
- a. id
  - b. first\_name
  - c. last\_name
  - d. date\_of\_birth
  - e. date\_of\_death
  - f. is\_visible
  - g. num\_related\_films
  - h. notes
  - i. created
  - j. created\_by
  - k. updated
  - l. Updated\_by
12. **Efis.efis\_company** - needs to be connected to country via id
- a. id
  - b. name
  - c. country\_id
  - d. type
  - e. year\_founded
  - f. ark\_code
  - g. is\_visible
  - h. num\_related\_films
  - i. notes
  - j. created
  - k. created\_by
  - l. updated
  - m. updated\_by
13. **Efis.efis\_film\_production** - extracted with company information, empty columns removed
- a. id
  - b. film\_id
  - c. classifier\_code
  - d. company\_id
  - e. people\_id
  - f. seq
  - g. created
  - h. created\_by
  - i. updated
  - j. Updated\_by
14. **Efis.efis\_company** - needs to be connected to country via id, empty columns removed
- a. id
  - b. name

- c. country\_id
- d. type
- e. year\_founded
- f. ark\_code
- g. is\_visible
- h. num\_related\_films
- i. notes
- j. created
- k. created\_by
- l. updated
- m. updated\_by

15. Efis.efis\_film\_purchase

- a. id
- b. film\_id
- c. type\_code
- d. company\_code
- e. url
- f. seq
- g. created
- h. created\_by
- i. updated
- j. updated\_by

16. Efis.efis\_film\_quote - DISCARD as only 8 results

- a. id
- b. film\_id
- c. language
- d. said\_by
- e. content
- f. seq
- g. created
- h. created\_by
- i. updated
- j. Updated\_by

17. Efis.efis\_film\_role\_person - these are roles played by a person in a film aka characters on screen - DISCARD as only 8 results

- a. id
- b. film\_id
- c. language
- d. classifier\_code
- e. people\_id
- f. role\_name
- g. seq
- h. created

- i. created\_by
- j. updated
- k. Updated\_by

18. Efis.efis\_film\_screening - extracted as is, needs to be connected to country via id

- a. id
- b. film\_id
- c. language
- d. classifier\_code
- e. country\_id
- f. location
- g. event\_info
- h. seq
- i. created
- j. created\_by
- k. updated
- l. Updated\_by

19. Efis.efis\_film\_screening\_result - DISCARD as only 8 results

- a. id
- b. film\_id
- c. language
- d. event\_info
- e. seq
- f. country\_id
- g. city
- h. visitor\_count
- i. visitors\_from
- j. visitors\_to
- k. created
- l. created\_by
- m. updated
- n. Updated\_by

20. Efis.efis\_film\_title\_episode - extracted as is, few entries - 237

- a. id
- b. film\_id
- c. language
- d. title
- e. created
- f. created\_by
- g. updated
- h. updated\_by

21. Efis.efis\_film\_title\_parallel - extracted as is

- a. id



- b. film\_id
- c. language\_code
- d. title
- e. created
- f. created\_by
- g. updated
- h. updated\_by

22. **Efis.efis\_film\_track** - extracted as is

- a. id
- b. film\_id
- c. language\_code
- d. classificator\_code
- e. seq
- f. created
- g. created\_by
- h. updated
- i. updated\_by

23. **Efis.efis\_film\_translation** - film titles in different languages - this table is two large and so is split into two during the extract - **nr 1** contains title information; **nr 2** contains content descriptive information. Nr 2 was extracted as a tab separated text file as Excel was not able to open it - perhaps Ksenia can work the file into a google sheet

- a. id
- b. film\_id
- c. language
- d. title\_proper
- e. title\_alternative
- f. title\_avant
- g. title\_working
- h. title\_serial
- i. title\_additional\_info
- j. era\_time
- k. content\_summary
- l. contents
- m. scenario
- n. bibliography
- o. makers\_remembrance
- p. film\_makers\_full\_list
- q. general\_material\_designation
- r. indoor\_shooting
- s. outdoor\_shooting
- t. trivia
- u. tagline
- v. short\_introduction

w. synopsis

- x. buyers\_info
- y. consumer\_carrier
- z. updated
- aa. updated\_by
- bb. created
- cc. created\_by

24. Efis.efis\_film\_keyword - connected to keyword values and keyword groups

- a. id
- b. film\_id
- c. keyword\_classifier\_code
- d. keyword\_id
- e. seq
- f. updated
- g. updated\_by
- h. created
- i. Created\_by

25. Efis.efis\_keyword\_translation

- m. id
- n. language
- o. keyword\_id
- p. value
- q. updated
- r. updated\_by
- s. Created
- t. Created\_by

26. Efis.efis\_keyword\_keyword\_group

- u. id
- v. keyword\_id
- w. keyword\_group\_id
- x. updated
- y. updated\_by
- z. created
- aa. created\_by

27. Efis.efis\_keyword\_group\_translation

- bb. id
- cc. language
- dd. keyword\_group\_id
- ee. value
- ff. updated
- gg. updated\_by
- hh. Created
- ii. Created\_by

28. Efis.efis\_keyword\_group

- jj. id

kk. type

ll. is\_visible

mm. updated

nn. updated\_by

oo. created

pp. created\_by

## Additions

29. Efis.Efis\_film\_festival - DISCARD as only 44 results

1. id,

2. language,

3. film\_id,

4. festival\_id,

5. festival\_year,

6. participation\_code,

7. participation\_relation,

8. seq,

9. created,

10. created\_by,

11. updated,

12. updated\_by

30. Efis.Efis\_festival - DISCARD as only 44 results

1. id,

2. original\_name,

3. link,

4. updated,

5. updated\_by,

6. created,

7. created\_by,

8. continent\_type\_code

31. Efis.Efis\_festival\_translation - DISCARD as only 44 results

1. id,

2. language,

3. festival\_id,

4. name,

5. updated,

6. updated\_by,

7. created,

8. created\_by

32. Efis.efis\_film\_physical\_description - extracted as is

1. id,

2. film\_id,

3. classifier\_type\_code,

4. classifier\_code1,

5. classifier\_code2,
6. classifier\_code3,
7. amount,
8. seq,
9. element\_number,
10. created,
11. created\_by,
12. updated,
13. updated\_by

33. Efis.efis\_film\_relation - extracted as is

1. id,
2. film\_id,
3. classifier\_code,
4. film\_seq\_related\_id,
5. seq,
6. created,
7. created\_by,
8. updated,
9. updated\_by

34. Efis.efis\_film\_seq - extracted as is

1. id,
2. scenario\_file\_bundle\_id,
3. role\_file\_bundle\_id,
4. physical\_file\_bundle\_id,
5. credits\_file\_bundle\_id,
6. music\_sheet\_file\_bundle\_id,
7. subtitle\_file\_bundle\_id

Not needed:

35. Efis.efis\_people\_image\_translation
36. Efis.efis\_people\_image

FURTHER RESEARCH:

Could also extract if needed, but the files are too large due to heaps textual information and queries are crashing - perhaps let's leave this for later if needed:

37. Efis.efis\_people\_description

- qq. id
- rr. people\_id
- ss. language
- tt. biography
- uu. introduction
- vv. award
- ww. bibliography

xx. information

yy. created

zz. created\_by

aaa. updated

bbb. updated\_by

38. efis.efis\_company\_description

ccc. id

ddd. company\_id

eee. language

fff. award

ggg. introduction

hhh. biography

iii. information

jjj. contact

kkk. created

lll. created\_by

mmm. updated

nnn. updated\_by

## Appendix 2 - MySQL queries used to extract data concerning newsreels from the EFDB collection

# extract newsreel IDs vol. 1

```
SELECT * FROM efis.efis_film_type_rel
WHERE film_type_id = "97" OR film_type_id = "96" OR film_type_id = "21" OR
film_type_id = "59" OR film_type_id = "110";
```

# extract newsreel IDs vol. 2

```
SELECT * FROM efis.efis_film_type_rel_veebist
WHERE film_type_id = "97" OR film_type_id = "96" OR film_type_id = "21" OR
film_type_id = "59" OR film_type_id = "110";
```

# a table with unique extracted newsreel ids is re-uploaded to the EFIS database as vz\_unique\_newsreel\_ids

# extract information needed from efis.efis\_film

```
SELECT vz_unique_newsreel_ids.film_id, film_seq_id, film_version_number,
production_year, is_production_year_assumed, archive_arrival_date, archive_fond_name,
archive_fond_nr, archive_archival_document_nr
FROM vz_unique_newsreel_ids
LEFT JOIN efis.efis_film
ON efis_film.id = vz_unique_newsreel_ids.film_id
```

# extract information needed from efis.efis\_film\_country

```
SELECT vz_unique_newsreel_ids.film_id, country_id, seq
FROM vz_unique_newsreel_ids
LEFT JOIN efis.efis_film_country
ON efis_film_country.film_id = vz_unique_newsreel_ids.film_id
```

# extract information on distributor needed from efis.efis\_film\_distributor, efis.efis\_company

```
SELECT vz_unique_newsreel_ids.film_id, language, classifier_code, company_id, seq,
name AS company_name, country_id, type, year_founded, num_related_films, notes
FROM vz_unique_newsreel_ids
LEFT JOIN efis.efis_film_distributor
ON efis_film_distributor.film_id = vz_unique_newsreel_ids.film_id
LEFT JOIN efis_company
ON efis_company.id = efis_film_distributor.company_id
```

# extract information needed from efis.efis\_film\_edition

```
SELECT vz_unique_newsreel_ids.film_id, lang, classifier_code, title, seq
FROM vz_unique_newsreel_ids
LEFT JOIN efis_film_edition
ON efis_film_edition.film_id = vz_unique_newsreel_ids.film_id
```

```
# extract information on cast (from efis.efis_film_maker, efis_people)
SELECT vz_unique_newsreel_ids.film_id, classificator_code, efis_film_maker.people_id, seq,
first_name, last_name, date_of_birth, date_of_death, num_related_films, notes
FROM vz_unique_newsreel_ids
LEFT JOIN efis_film_maker
ON efis_film_maker.film_id = vz_unique_newsreel_ids.film_id
LEFT JOIN efis_people
ON efis_people.id = efis_film_maker.people_id
```

```
# extract information on film music and music makers + companies (from efis_film_music,
efis_film_music_maker, efis_people, efis_company)
SELECT vz_unique_newsreel_ids.film_id, efis_film_music.id as film_music_id,
efis_film_music.title as film_music_title, efis_film_music.seq as film_music_seq,
efis_film_music_maker.classificator_code as music_maker_classificator_code,
efis_film_music_maker.company_id, efis_film_music_maker.people_id,
efis_film_music_maker.seq as film_music_maker_seq,
first_name, last_name, date_of_birth, date_of_death, efis_people.num_related_films as
people_num_related_films, efis_people.notes as people_notes,
efis_company.name as company_name, efis_company.country_id, efis_company.type as
company_type, year_founded, efis_company.num_related_films as
company_num_related_films, efis_company.notes as company_notes
FROM vz_unique_newsreel_ids
LEFT JOIN efis_film_music
ON efis_film_music.film_id = vz_unique_newsreel_ids.film_id
LEFT JOIN efis_film_music_maker
ON efis_film_music_maker.film_music_id = efis_film_music.id
LEFT JOIN efis_people
ON efis_people.id = efis_film_music_maker.people_id
LEFT JOIN efis_company
ON efis_company.id = efis_film_music_maker.company_id
```

```
# extract information on film production companies from efis.efis_film_production,
efis.efis_company
SELECT vz_unique_newsreel_ids.film_id, classificator_code,
efis_film_production.company_id, seq,
efis_company.name AS company_name, efis_company.country_id, efis_company.type,
year_founded, efis_company.num_related_films, efis_company.notes
FROM vz_unique_newsreel_ids
LEFT JOIN efis_film_production
ON efis_film_production.film_id = vz_unique_newsreel_ids.film_id
LEFT JOIN efis_company
ON efis_company.id = efis_film_production.company_id
```

```
# DISCARD AS ONLY 8 ENTRIES extract information needed from efis.efis_film_quote
SELECT efis_film_quote.film_id, language, said_by, content, seq
```

```
FROM efis_film_quote, vz_unique_newsreel_ids
WHERE efis_film_quote.film_id = vz_unique_newsreel_ids.film_id
```

```
# DISCARD AS ONLY 14 ENTRIES extract information needed from
efis.efis_film_role_person
SELECT efis_film_role_person.film_id, language, classificator_code, people_id, role_name,
seq
FROM efis_film_role_person, vz_unique_newsreel_ids
WHERE efis_film_role_person.film_id = vz_unique_newsreel_ids.film_id
```

```
# extract information needed from efis.efis_film_screening
SELECT vz_unique_newsreel_ids.film_id, language, classificator_code, country_id, location,
event_info, seq
FROM vz_unique_newsreel_ids
LEFT JOIN efis_film_screening
ON efis_film_screening.film_id = vz_unique_newsreel_ids.film_id
```

```
# DISCARD AS 0 ENTRIES extract information needed from efis.efis_film_screening_result
SELECT id, efis_film_screening_result.film_id, language, event_info, seq, country_id, city,
visitor_count, visitors_from, visitors_to
FROM efis_film_screening_result, vz_unique_newsreel_ids
WHERE efis_film_screening_result.film_id = vz_unique_newsreel_ids.film_id
```

```
# extract information needed from Efis.efis_film_title_episode
SELECT vz_unique_newsreel_ids.film_id, language, title
FROM vz_unique_newsreel_ids
LEFT JOIN efis_film_title_episode
ON efis_film_title_episode.film_id = vz_unique_newsreel_ids.film_id
```

```
# extract information needed from Efis.efis_film_title_parallel
SELECT vz_unique_newsreel_ids.film_id, language_code, title
FROM vz_unique_newsreel_ids
LEFT JOIN efis_film_title_parallel
ON efis_film_title_parallel.film_id = vz_unique_newsreel_ids.film_id
```

```
# extract information needed from Efis.efis_film_track
SELECT vz_unique_newsreel_ids.film_id, language_code, classificator_code, seq
FROM vz_unique_newsreel_ids
LEFT JOIN efis_film_track
ON efis_film_track.film_id = vz_unique_newsreel_ids.film_id
```

```
# extract information needed from Efis.efis_film_translation - part1: title info
SELECT vz_unique_newsreel_ids.film_id, language, title_proper, title_alternative, title_avant,
title_working, title_serial, title_additional_info
FROM vz_unique_newsreel_ids
LEFT JOIN efis_film_translation
```



```
ON efis_film_translation.film_id = vz_unique_newsreel_ids.film_id
```

```
# extract information needed from Efis.efis_film_translation - part2: description info -  
extracted as a text file
```

```
SELECT vz_unique_newsreel_ids.film_id, language, era_time, content_summary, contents,  
bibliography, film_makers_full_list, general_material_designation, trivia, short_introduction,  
synopsis  
FROM vz_unique_newsreel_ids  
LEFT JOIN efis_film_translation  
ON efis_film_translation.film_id = vz_unique_newsreel_ids.film_id
```

```
# extracting information on keywords from efis.efis_film_keyword,  
Efis.efis_keyword_translation, efis_keyword_keyword_group,  
efis_keyword_group_translation, efis_keyword_group  
SELECT vz_unique_newsreel_ids.film_id, keyword_classificator_code,  
efis_film_keyword.keyword_id, seq as keyword_seq,  
efis_keyword_translation.language as keyword_language, efis_keyword_translation.value as  
keyword_value,  
efis_keyword_keyword_group.keyword_group_id,  
efis_keyword_group_translation.language as keyword_group_language,  
efis_keyword_group_translation.value as keyword_group_value,  
efis_keyword_group.type as keyword_group_type  
FROM vz_unique_newsreel_ids  
LEFT JOIN efis_film_keyword  
ON vz_unique_newsreel_ids.film_id = efis_film_keyword.film_id  
LEFT JOIN efis_keyword_translation  
ON efis_film_keyword.keyword_id = efis_keyword_translation.keyword_id  
LEFT JOIN efis_keyword_keyword_group  
ON efis_keyword_translation.keyword_id = efis_keyword_keyword_group.keyword_id  
LEFT JOIN efis_keyword_group_translation  
ON efis_keyword_keyword_group.keyword_group_id =  
efis_keyword_group_translation.keyword_group_id  
LEFT JOIN efis_keyword_group  
ON efis_keyword_group_translation.keyword_group_id = efis_keyword_group.id
```

```
# DISCARD AS 44 ENTRIES extract information needed from Efis.Efis_film_festival  
SELECT vz_unique_newsreel_ids.film_id, language, festival_id, festival_year,  
participation_code, participation_relation, seq  
FROM efis_film_festival, vz_unique_newsreel_ids  
WHERE efis_film_festival.film_id = vz_unique_newsreel_ids.film_id
```

```
# extract information needed from Efis.efis_film_physical_description  
SELECT vz_unique_newsreel_ids.film_id, classificator_type_code, classificator_code1,  
classificator_code2, classificator_code3, amount, seq, element_number  
FROM vz_unique_newsreel_ids  
LEFT JOIN efis_film_physical_description
```

```
ON efis_film_physical_description.film_id = vz_unique_newsreel_ids.film_id
```

```
# extract information needed from Efis.efis_film_relation
```

```
SELECT vz_unique_newsreel_ids.film_id, classifier_code, film_seq_related_id, seq
```

```
FROM vz_unique_newsreel_ids
```

```
LEFT JOIN efis_film_relation
```

```
ON efis_film_relation.film_id = vz_unique_newsreel_ids.film_id
```

```
# extract information needed from Efis.efis_film_seq
```

```
SELECT vz_unique_newsreel_ids.film_id, efis_film.film_seq_id, scenario_file_bundle_id,
```

```
role_file_bundle_id, physical_file_bundle_id, credits_file_bundle_id,
```

```
music_sheet_file_bundle_id, subtitle_file_bundle_id
```

```
FROM vz_unique_newsreel_ids
```

```
LEFT JOIN efis.efis_film
```

```
ON efis_film.id = vz_unique_newsreel_ids.film_id
```

```
LEFT JOIN efis_film_seq
```

```
ON efis_film.film_seq_id = efis_film_seq.id
```