MINING the MEDIATIZED MEMORY

#DHH16

Digital
Humanities
Hackathon

Anu Koivunen
Emily Öhman
Julia Poutanen
Matti Lassila
Salla-Maaria Laaksonen
Sezin Yaman

DHH16
DIGITAL HUMANITIES HACKATHON
MAY 2016 HELSINKI

Research Frame

Investigating YLE programming of 1957-1990, we study "an era of scarcity": one or two TV channels, broadcasting for part of the day only, operating as a "powerful instrument of social integration" and structuring, through scheduling, "the nation's private life" (Ellis 2000). Since the history of YLE has mostly been written from the perspective of the organization (Lyytinen & Vihavainen 1996, Salokangas 1996) or case studies of programme genres, the metadata offers a unique opportunity to study the whole of programming: to map the proportions of different programme categories and to track the topics defining "the mediatized centre" (Couldry 2003), an index of the national public sphere.

While ultimately seeking to understand whether Finnish TV history features the dynamics of everyday television and media events (Dayan & Katz 2002, Couldry, Hepp & Krotz 2010) as well as the politics of the TV flow (representation of persons, parties, genders, topics), the hackathon ended up being an exercise in historiography: a study of the dataset and its allowances.

Data Refining and Categorization

» We use the metadata of **YLE Metro TV archive collection between 1.1.1957–31.12.1990**, in total 96 838 records, feature id-number, titles, contributors, various description and content fields, subtitles, duration and classifications for each programme.

» The original XML files were transformed to csv. Refining the data and classifying the program types was done using **search queries in OpenRefine by four coders**. We used the programme categories of the Ministry of Transport and Communications: news (UUTI), actualities (AJAN), factual programming (ASIA), feature films (ELOK), domestic and foreign fiction (KFIK/UFIK), children's programmes (LAST), sports (URHE), lifestyle (LIFE), reality (REAL), and other.

» Currently **79 670 of 96 839 entries have been categorized**, but there are identification errors and interpretation differences. A consistent **validation** procedure is needed.

» No clear distinction between actualities and factual programmes, the latter ending up being as a massive category. Also the distinction between cultural programmes and entertainment is fluid. There is no separate category for religious programme, now categorized as lifestyle, and none for various media events (parades etc.) beyond sports events.
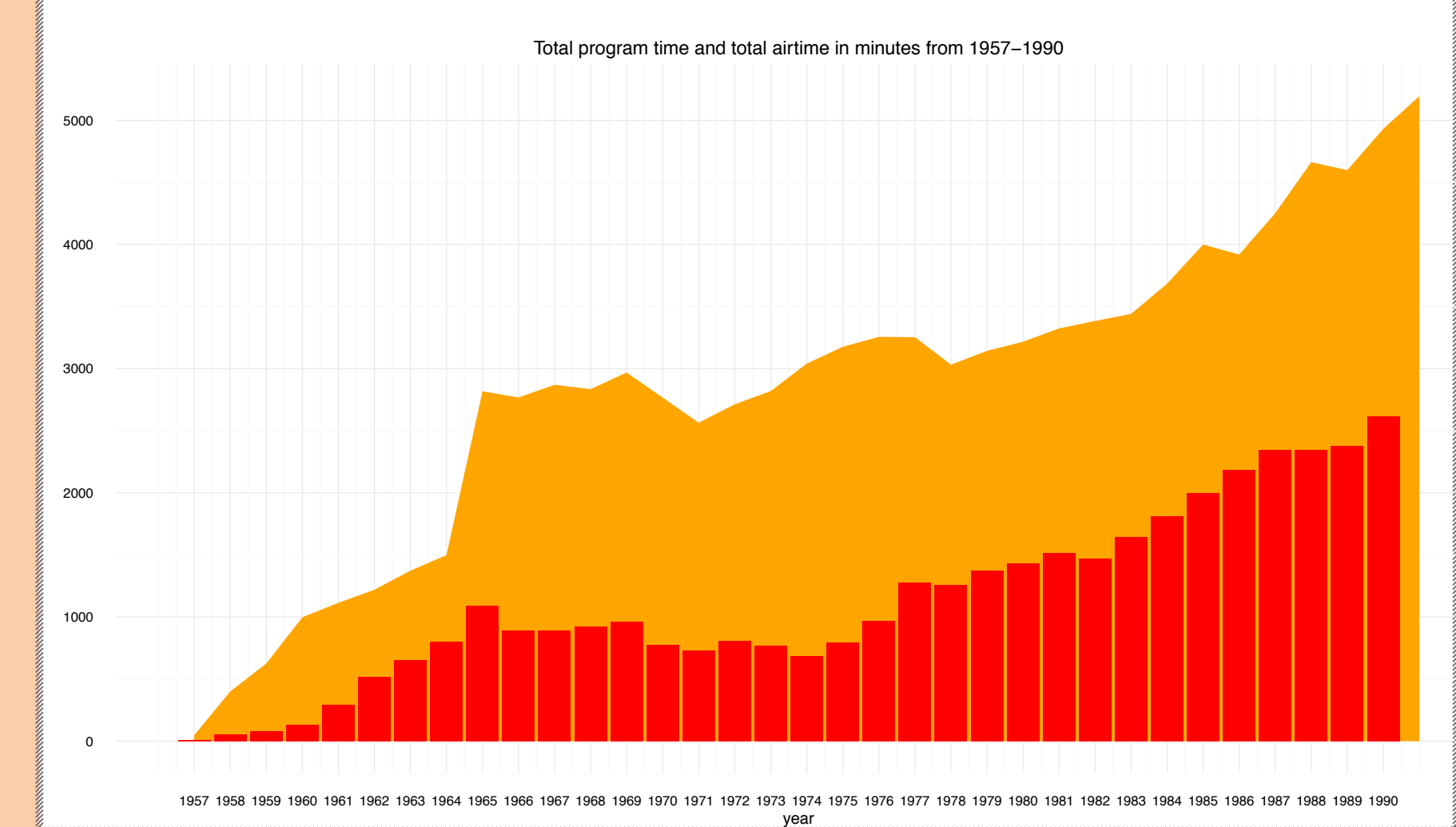
Data Context

» Our data is **metadata**, i.e. it does not include the original TV programs but fields that describe their properties and content. These fields have been filled in by tens of different archivists.

» Only **selected programs were archived until 1984**, live programming is mosty excluded. During 1977-1984 roughly half of the program tapes were reused. Hence, many original recordings are lost and cannot be accessed for retrospective archiving.

» **MTV programmes are not featured**, despite the fact that it operated under YLE licence and rented airtime in YLE channels 1 and 2 before channel 3 was launched in 1987 as a fully commercial national broadcasting channel.

» To assess the overall significance of topic analysis, we would need **comparative datasets** (e.g. newspapers) and preferably MTV3 data.

Total program time and total airtime in minutes from 1957–1990

Programme type by year (%)
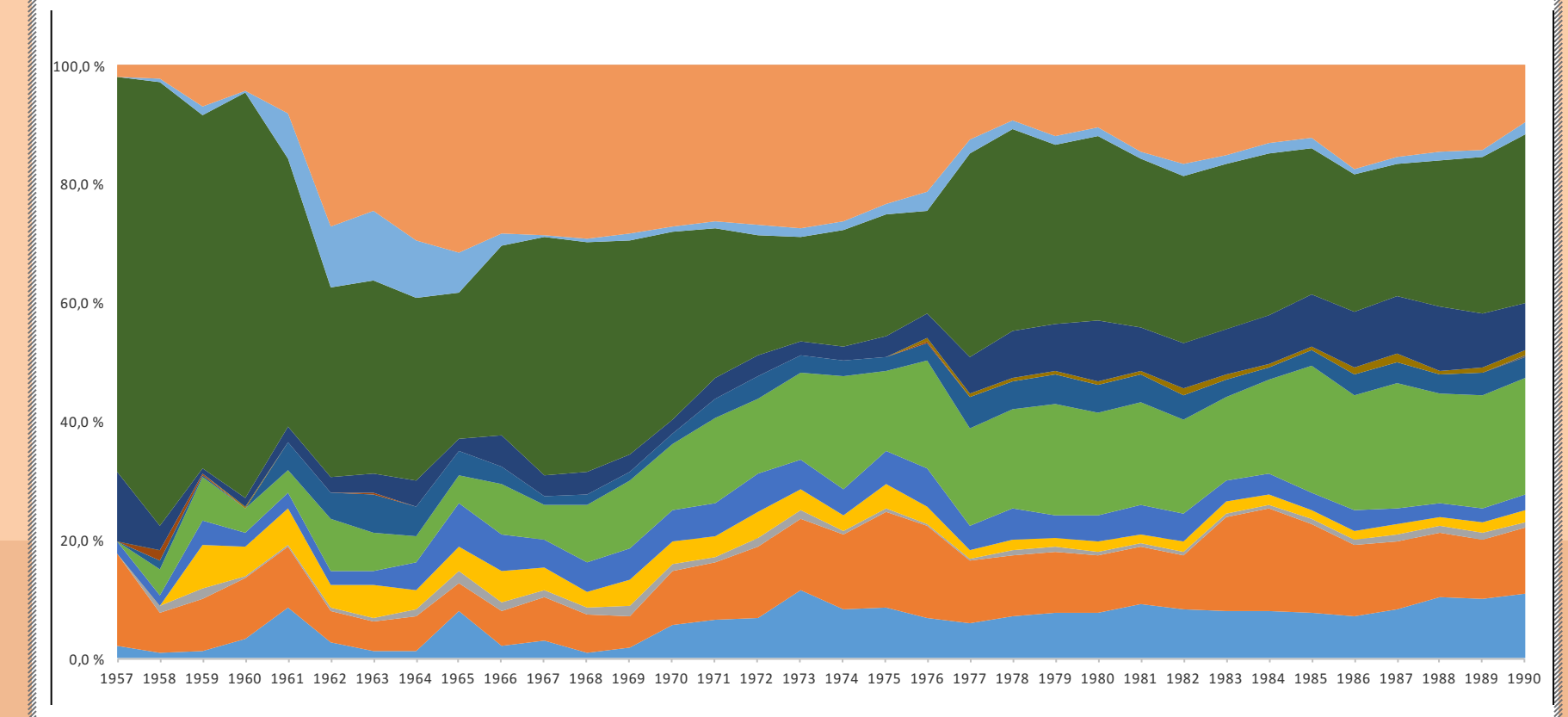
Next Steps

» **Structural Topic Modelling** will be used to identify the recurrent topics in the programme descriptions. STM allows the use of all the variables in data in generating the topics.

» **Named Entity Recognition** will be used to identify all named entities, such as persons, places, and organizations in the data. This allows us to investigate the share of airtime for different actors as well as. e.g., different genders.