

# Metadata in CHAOS

## How researchers tag and annotate radio broadcasts

Marianne Lykke, Aalborg University

Haakon Lund, RSLIS, Copenhagen University

Mette Skov, Aalborg University

  
e-Learning Lab

Center for User Driven Innovation, Learning and Design

# LARM AUDIO RESEARCH ARCHIVE

- LARM Audio Research Archive
  - Digital research infrastructure for radio and auditory cultural heritage, 2010-2014
  - Funded by the Danish Ministry of Science, Technology and Innovation
- A consortium of 13 Danish research institutions and archives
  - University partners 
  - Danish Broadcast Corporation
  - The State and University Library
  - Danish Research Network
  - The Danish Media Museum
  - Roskilde Museum of Contemporary Art

Royal School of Library and  
Information Science  
University of Copenhagen  
Roskilde University  
University of Aarhus  
University of Aalborg  
University of Southern Denmark

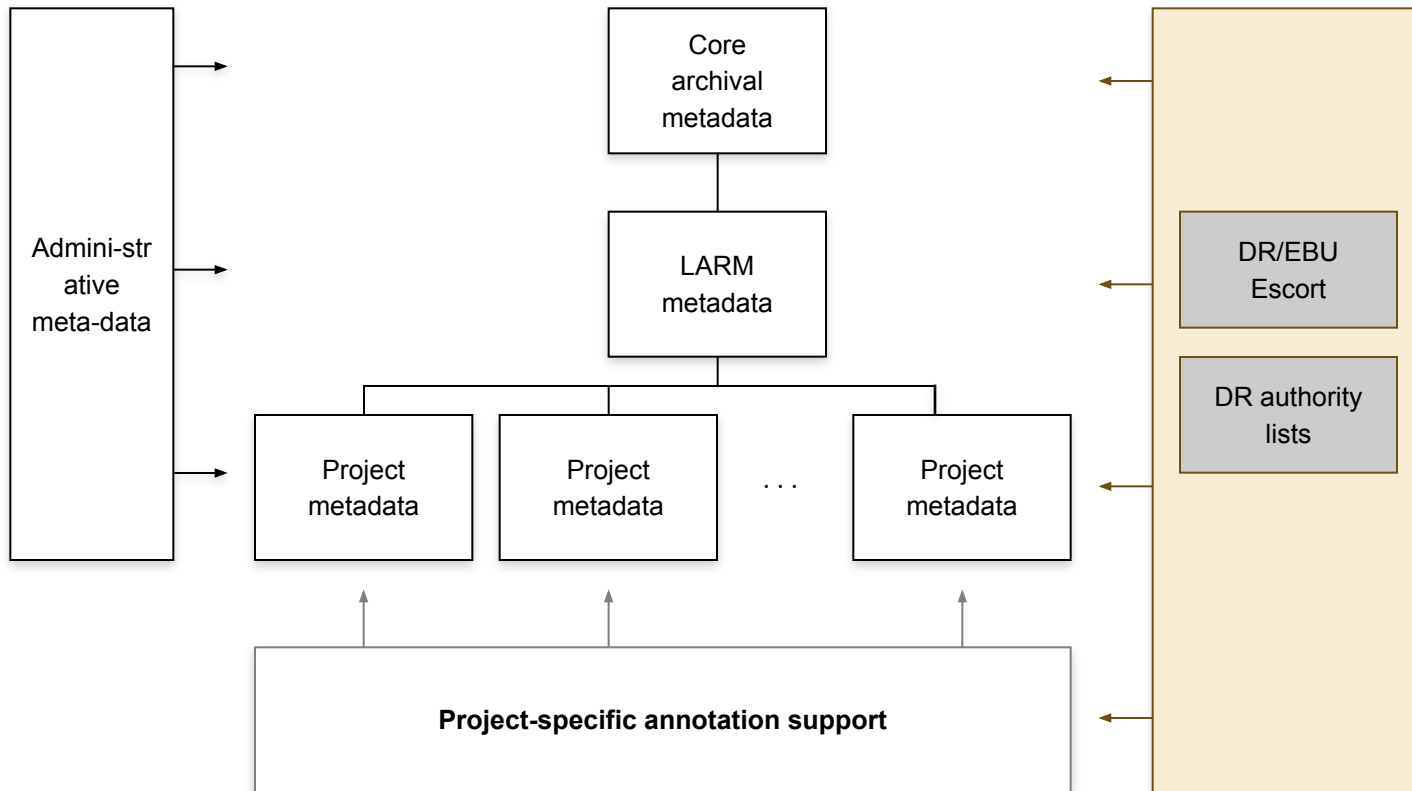
# Access to radio broadcast



- Challenges:
  - Amount of data, 1+ mill. programme hours
  - Access, technical and legal issues
  - Different archives with different principles for selection and metadata
  - How to search for sound?
- Focus of the present study: how researchers apply the metadata scheme in their research work
- Earlier studies show interesting results on user tagging in the cultural heritage domain (Hollink et al. 2004, Gligorov et al. 2011, Oomen et al. 2010)



# Implemented metadata schema



(See Lund et al. 2013 and Skov et al. 2012)

# Research question and design

- RQ: How and with what purpose do LARM researchers apply metadata and annotations to radio broadcasts as part of their research work?
- A practice-oriented study focusing on researchers' incorporation, adoption, and adaption of metadata into working practice (Dourish 2003).
- The analysis is based on two studies:
  - A qualitative study of subjects and vocabulary of the applied metadata and annotations and,
  - 5 semi-structured interviews with researchers about their goals for tagging.

# Example of a LARM metadata description in Larm.fm

The screenshot shows the Larm.fm web interface. The browser's address bar displays the URL 'Lille husmor, hvad nu? Hørebillede om f...'. The page title is 'Lille husmor, hvad nu? Hørebillede om f...'. The main content area is titled 'Lille husmor, hvad nu? Hørebillede om f...' and features a play button icon. Below the play button, the following metadata is displayed:

Tidspunkt	04. jan. 1957 kl. 12:00
Varighed	54:40
Kanal	Ukendt

Below the metadata, there are buttons for 'Eksportér XML', 'Rediger', and 'Danish +'. The 'LARM-metadata' section contains the following text:

Program titel -Hvis programtitel i arkivmetadata ikke er fyldestgørende:

Beskrivelse:

"De unge piger foretrækker uddannelse frem for udstyr". I programmet bliver nogle unge piger interviewet om deres drømme og ønsker til uddannelse. Kvindelige fabriksarbejdere fortæller om både at have lønarbejde og passe hjem og børn.

Viggo Clausen, tilrettelægger / Eva Ree Hinrichsen, programmedarbejder

Genre:

Emne:

The interface also shows a sidebar with a 'Skema' menu and a 'Vedhæftede filer' section with a 'Vedhæft fil her' button.

# Findings: LARM metadata

- Metadata assigned to 480 broadcasts
- Most frequently to Title, followed by Tag, Genre, Description, and Subject
- Many metadata is often **placed incorrectly**
- Researchers aim at **providing neutral, descriptive** LARM metadata
- High level of **trust** in metadata provided by other users



# Findings: Project metadata and annotations

- Project metadata are primarily **content metadata**
- The Project metadata have several forms, e.g. single terms, transcriptions, citations, long and short annotations, and annotations in form of headlines.
- Applying annotations is **intertwined in the individual scholars research process**
- Annotations can **support the different steps** in the research process
- The researchers mainly view their annotations as **descriptive** and aim at **objectivity**.

# Types of annotations



Types of annotations	Examples of annotations
Indicative summary	Introduction to broadcast. Mr. Lindum says good evening in English
Informative summary	Talked about his adventures during the 1864 war. Talked again about his missionary work.
Analytical summary	WW2: Wives of Danish soldiers invited by Himmler to visit husbands in Germany.
Sound elements	Car. Door slam. Driving. Speaker voice.
Transcription	Ladies and gentlemen. A new year begins and we leave the old year without sorrow. (...)
Incomplete or inaudible broadcast	Error in sound file (loud noise)
Reserch related information	Various test data
Formal composition	Jingle: separator #2 followed by separator #1.

# Conclusion

- Tagging in LARM.fm is not collaborative. Instead assigned metadata are **individual** and **research specific**
- The main purpose is to administer **personal scientific analysis**
- A secondary purpose is to apply more general metadata for the benefit of the community
- The primary role of the LARM.fm archive is to provide **access** to broadcasts and provide **tools**
- **Knowledge organization tools and methods** can continuously play an important role in improving the research infrastructure

# LARM AUDIO RESEARCH ARCHIVE

Thank you for listening

Questions?

Read more about LARM: <http://www.larm-archive.org/about-larm/>

# References

- Dourish, P. 2003. The appropriation of interactive technologies. *Computer Supported Cooperative Work*, 12, 465-490.
- Gligorov, R., Hildebrand, M., van Ossenbruggen, J., Schreiber, G., and Aroyo, L. 2011. On the role of user-generated metadata in audio visual collections. In *Proceedings of the sixth International Conference on Knowledge Capture* (pp. 145-152). ACM.
- Hollink, L., Schreiber, A. T., Wielinga, B. J., and Worring, M. 2004. Classification of user image descriptions. *International Journal of Human-Computer Studies*, 61(5), 601-626.
- Lund, H., Larsen, B., Bogers, T. and Lykke, M. 2013. CHAOS: User-driven development of metadata scheme for radio broadcast archives. *Proceedings of the iConference 2013* (pp. 990-994). IDEALS: iSchools.
- Oomen, J., Baltussen, L.B., Limonard, S., van Ees, A., Brinkerink, M., Aroyo, L., Vervaart, J., Asaf, K., and Gligorov, R. 2010. Emerging practices in the cultural heritage domain: social tagging of audiovisual heritage. In *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*.
- Skov, M. and Lykke, M. 2012. Unlocking radio broadcasts: user needs in sound retrieval. *Proceedings of the 4<sup>th</sup> International conference on Information Interaction in Context* (pp. 298-301). ACM.