

Copyrighted Media for Research?

Yle, MeMAD Project and Future Plans



MeMAD project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780069. This presentation has been produced by the MeMAD project. The content in this presentation represents the views of the authors, and the European Commission has no liability in respect of the content.





MeMAD

Methods for Managing
Audiovisual Data



memad.eu
info@memad.eu

 @memadproject
 MeMAD Project

What data would you need?



MeMAD project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780069. This presentation has been produced by theMeMAD project. The content in this presentation represents the views of the authors, and the European Commission has no liability in respect of the content.

MeMAD

Research and innovation project 2018-2020: multimodal content analysis and language technologies

- ASR
- Video and sound description
- Machine translation

Combining automated methods together and with human work

Domain: media industries

- media distribution (incl. broadcasting)
- media production
- translation, subtitling, audio descriptions



MeMAD project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780069. This presentation has been produced by the MeMAD project. The content in this presentation represents the views of the authors, and the European Commission has no liability in respect of the content.



MeMAD
Methods for Managing
Audiovisual Data

MeMAD consortium

Aalto University, Finland (Coordinator)

University of Helsinki, Finland

EURECOM, France

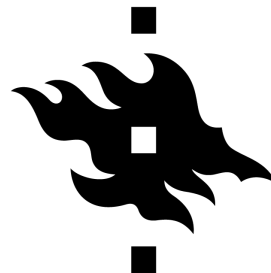
University of Surrey

Yle, the Finnish Broadcasting Company

Limecraft, Belgium

Lingsoft, Finland

Institut national audiovisuel, France



MeMAD project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780069. This presentation has been produced by the MeMAD project. The content in this presentation represents the views of the authors, and the European Commission has no liability in respect of the content.



MeMAD
Methods for Managing
Audiovisual Data

Evaluation through prototyping

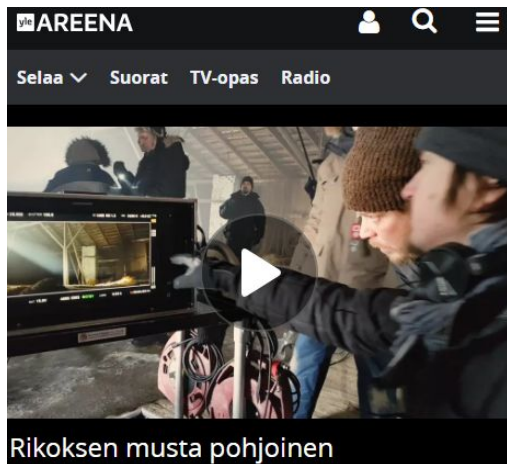
- Prototype platform combines the project's technologies
- New version each year
- Put to test by media professionals at Yle and external partners

Key questions:

- Where should we use automation?
- How to combine human work with automated analyses?



More information by automation



Audio, additional languages and audio description

Subtitles, additional languages

ASR, NER..

Structural analysis, themes, narration..

Visual analysis, faces, objects, actions..

Content descriptions: topics, subjects, visual content etc.

Semantic linking to other parts,
programs and resources



MeMAD
Methods for Managing
Audiovisual Data

Datasets fueling research

1. Research and open access datasets: TRECVID, MS-COCO etc.

- a. standard reference points for research groups
- b. generic annotation available, but possibly biased

2. Proprietary datasets: INA, Yle

- a. **existing contracts and rights issues affect availability**
- b. real-world examples of the targeted domain
- c. real-world annotations [both good and bad]



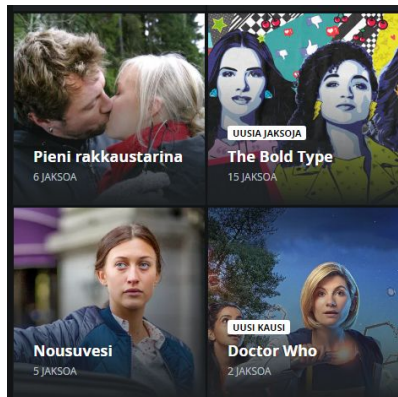
MeMAD

Methods for Managing
Audiovisual Data

Right to do what?

Copyright, creators
directors, writers,
actors, composers etc

**Commercial
contracts**
production,
distribution etc.



Typically media companies and archives **have some, but not all rights** to the content they have.

Additional rights can be negotiated and acquired, if needed.



Extended collective licensing

- Extended collective licensing [[Sopimuslisenssi](#)]:
 - A copyright society may grant licenses on behalf of all rights holders for some specific purposes
 - One contract instead of, say, 2500 individual contracts with each rights holder
 - E.g. making copies for research use
- Long term availability instead of single project datasets and licenses?
 - FAIR-principles, open science etc.



Talking helps!

Do not be afraid of lawyers and copyright. Focus on finding and negotiating solutions.

This is still new to copyright societies and rights holders. Explain what you wish to do.

Copyrighted content is available for research, if you need it.



MeMAD
Methods for Managing
Audiovisual Data



Future datasets?

[What data would you need?]





Yle wishes to open datasets

What is needed the most at the moment?

Smart subsets from these could be selected and shared to power the ecosystem.

Content: Video, audio, articles

400 000 video programs,
2,2 M audio programs,
400 000 images,
1 M articles
(estimated counts)

Other data

Subtitles,
content descriptions,
content tagging,
usage data,
past broadcast schedules,
program metadata,
other ?



MeMAD
Methods for Managing
Audiovisual Data



Where to start?

Maybe like this:

1-3 pilot projects: research, development, start ups

First versions of future framework license for Yle datasets could be based on these.

Interested?

Contact lauri.saarikoski@yle.fi

Thank you!

Text (+ annotations / metadata)?

Video (+ annotations / metadata)?

Audio (+ manuscripts / transcripts)?

Something completely different?



MeMAD

Methods for Managing
Audiovisual Data