EFG´s Approach to Authority File Building

SESSION 2: AUTHORITY FILES

EFG Workshop on Data Quality and Semantic Interoperability Issues in European Film Archives

Frankfurt am Main, 30 May 2011

Francesca Schulze
(Deutsches Filminstitut – DIF)
f.schulze@deutsches-filminstitut.de
www.europeanfilmgateway.eu

Cofunded by the Community Programme eContent*plus*

Overview of presentation

- Challenges (unifying names and titles)
- EFG metadata schema (filmographic part)
- Authority File Manager
- EFG portal example (role of filmographic data)
- Achievements
- Lessons learnt
- Visions: What could be done after EFG I?

Aggregation aims

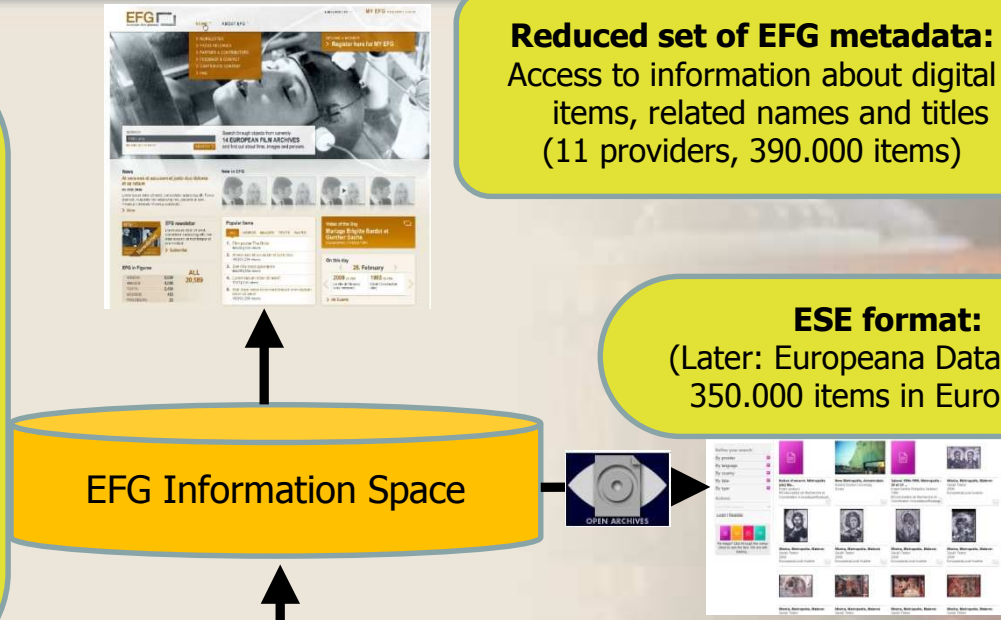**EFG metadata schema**
(925.000 XML records):

Aggregation of filmographic and biographic data as well as data on digital archival resources

2nd idea of EFG: Create a common registry of film works and persons active in the film domain => Approach: unique EFG authority files

Filmographic data in EFG: 120.000 film works, 240.000 persons, 26.000 corporations, + groups, + events

Reduced set of EFG metadata:
Access to information about digital items, related names and titles
(11 providers, 390.000 items)

ESE format:
(Later: Europeana Data Model)
350.000 items in Europeana

EFG Information Space

OPEN ARCHIVES

Metadata

Metadata

Metadata

Metadata

Ca. 13 more sites
(16 Providers)

Danish Filmography (DFI)

Filmportal.de (DIF)

Luce site

Challenges

1) **Doublets:** Information about the same person or film work can come from different data sources due to numerous European co-productions

2) **Diverse quality of metadata** on persons, film works and corporations delivered by partner archives:
   - Authority records
     - not all partners manage them
     - 14 partners delivered unique records for film works, 8 for persons, 5 for corporations
     - Name and title data came normalised and not normalised
   - Cataloguing practices
     - no common cataloguing rules applied at partner archives
     - records varying in level of detail on information provided
     - no archive references its names and titles to an external authorit file

Challenges

- Metadata structures
  - ➢ no common formats to deliver information on names and titles
  - ➢ most partners use in-house or an EFG adapted format
  - ➢ only 1 partner supplied this kind of information in a standardized format (EAC)

3) Data management issues: Activities to match & merge name and title data to EFG authority files must be in line with the ingestion, data cleaning and enrichment activites
   - Clean doublets in local and EFG database
   - Enrich name and title records (locally and in EFG)
   - Develop semantic technology tools (EFG backend tools)
   - Establish reliable ingestion and metadata editing workflow

Aggregation tools

➢ EFG metadata schema:

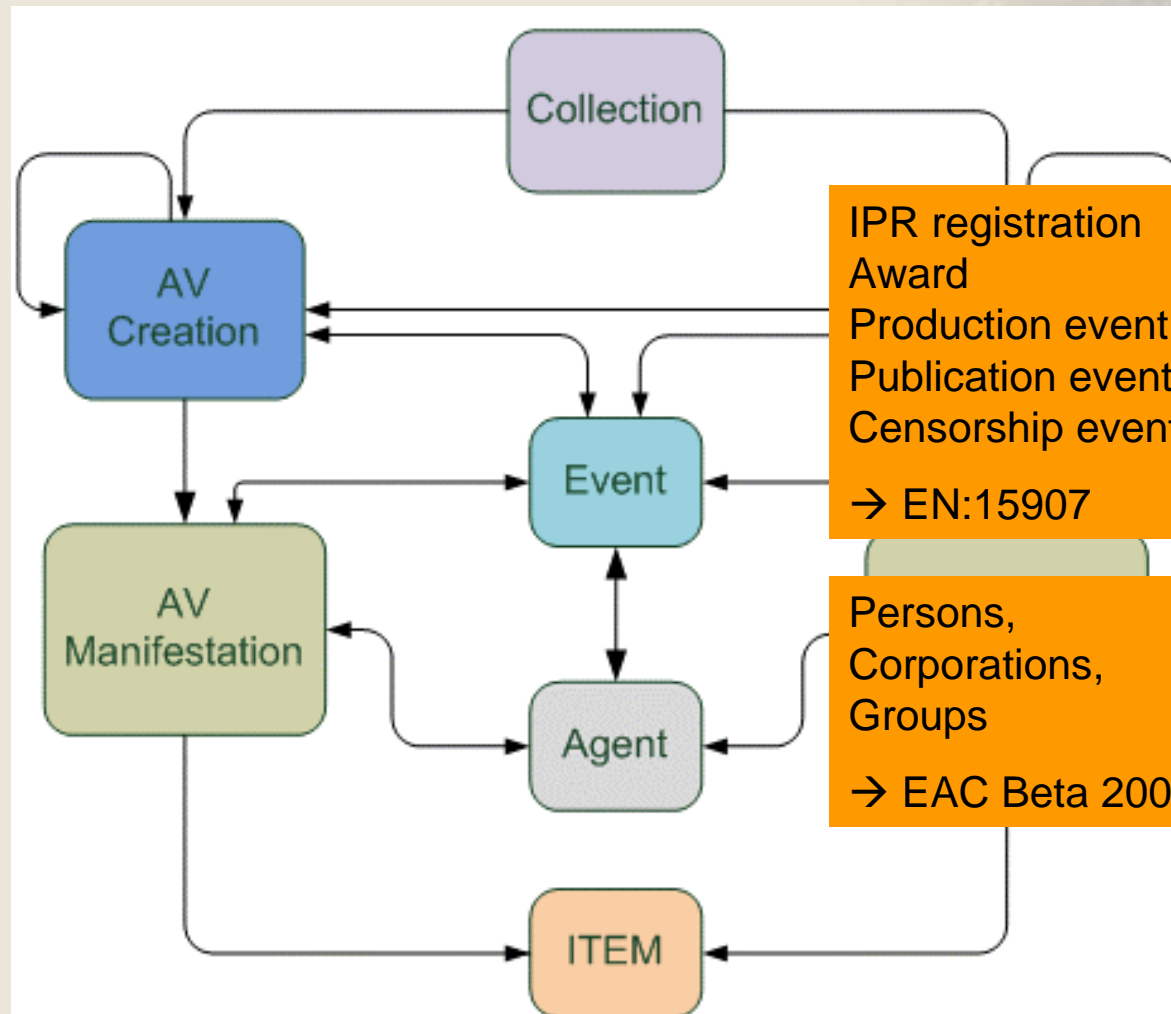Complex schema based on FRBR oriented Cinematographic Works Standard (EN:15907)

www.europeanfilmgateway.eu/guidelines_and_standards.php

➢ Authority File Manager (AFM):

Tool to identify doublet records of the same film work or person in EFG database

http://www.europeana.eu/portal/thoughtlab_improvingmetadata.html

EFG metadata schema

Film work

→ EN:15907

Film manifestation

→ EN:15907

Collection

AV Creation
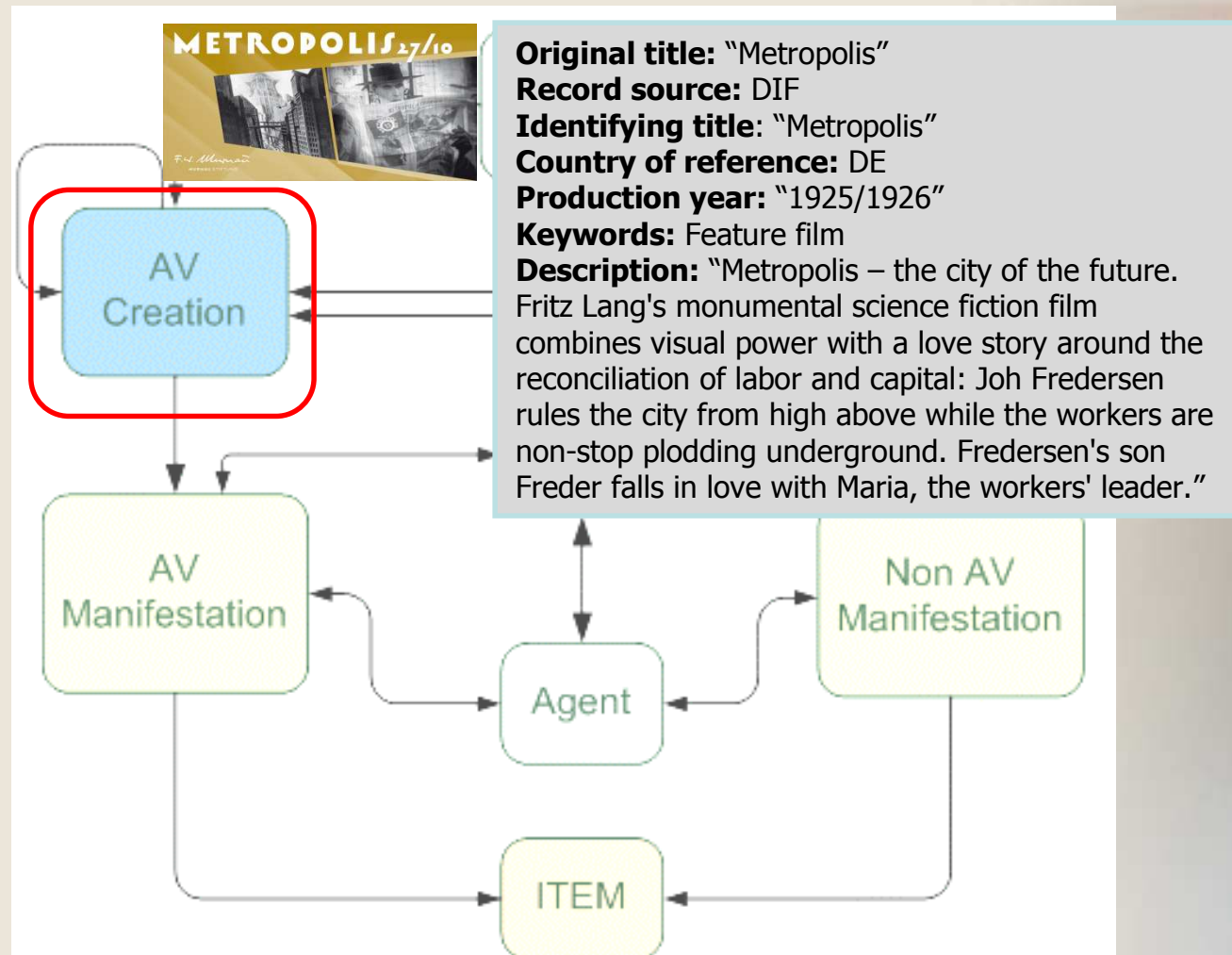
Event
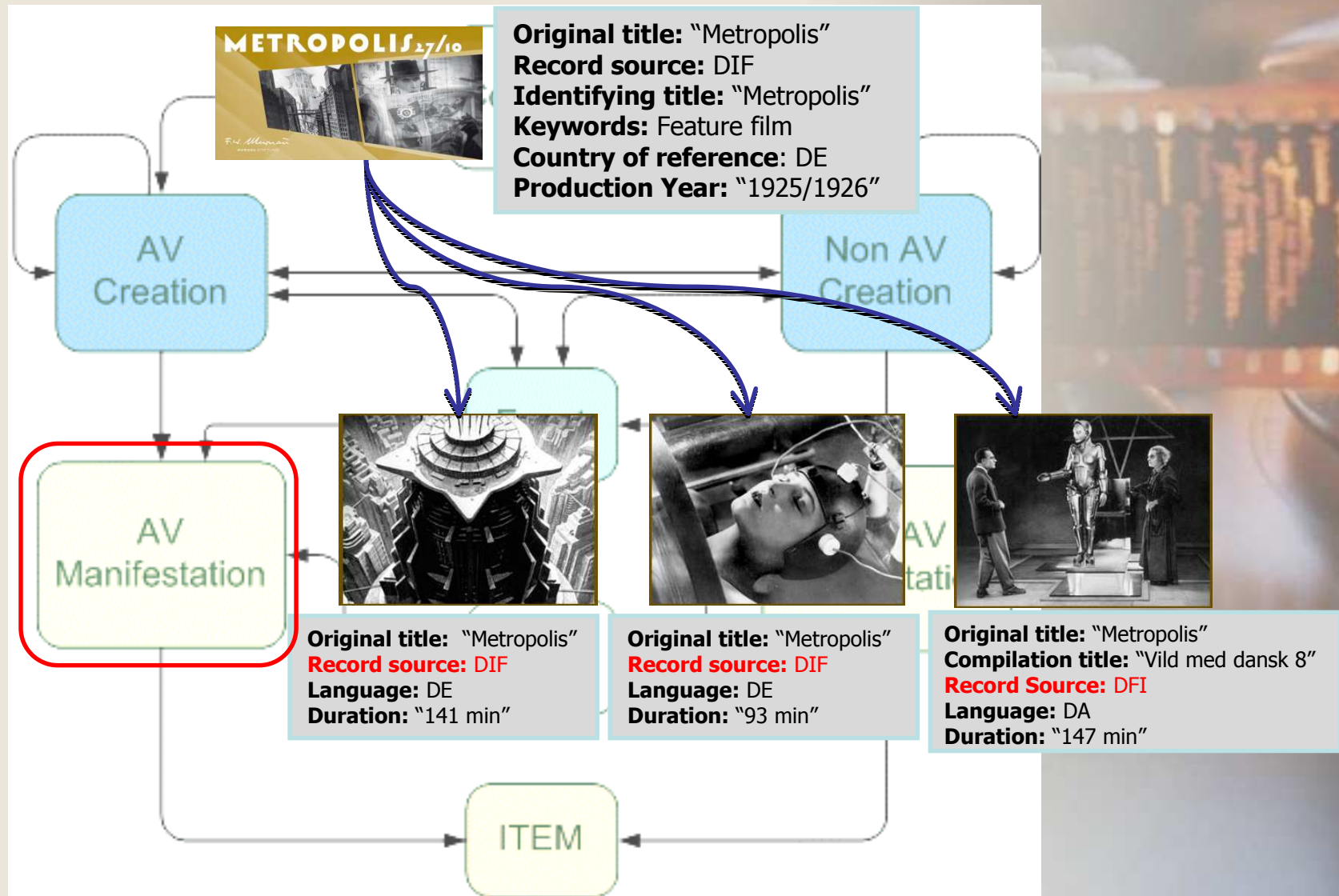
AV Manifestation

Agent

ITEM

IPR registration
Award
Production event
Publication event
Censorship event

→ EN:15907

Persons,
Corporations,
Groups

→ EAC Beta 2004

An example of the use of the EFG schema

Manifestations from different sources

Original title: "Metropolis"
Record source: DIF
Identifying title: "Metropolis"
Keywords: Feature film
Country of reference: DE
Production Year: "1925/1926"

AV Creation

Non AV Creation

AV Manifestation

Original title: "Metropolis"
Record source: DIF
Language: DE
Duration: "141 min"

Original title: "Metropolis"
Record source: DIF
Language: DE
Duration: "93 min"

Original title: "Metropolis"
Compilation title: "Vild med dansk 8"
Record Source: DFI
Language: DA
Duration: "147 min"

ITEM

EFG Backend Tools

➢ Vocabulary Checker

➢ Authority File Manager

➢ Metadata Editor

Authority File Manager

➢ Data cleaning in EFG

➢ Identifies doublet records of the same film work or person (active in the film domain) in EFG Information Space

• Allows partners also to detect doublets in their local databases

• Allows to merge doublet records to authority records

• Uses basic concepts of „Match & Merge" applied by librarians

Authority File Manager

Authority File Manager

EFG - Authority File Management

Person Record Merge

Back to detail page   Help

Choose which record to keep and select which version of the field value you want to preserve:

First name	Eva Maria		First name	Eva-Maria
Last name	Meineke		Last name	Meinecke
Date	1923-10-08		Date	
Type of activity	Actor		Type of activity	Actor
Sex	f		Sex	f
Place of birth	Berlin		Place of birth	
Provider	Deutsches Filminstitut - DIF		Provider	Deutsches Filminstitut - DIF
Identifier:	person.filmportal.de/DIF_person_7E2201C16!		Identifier:	person.filmportal.de/DIF_person_F909E564D

Switch to this record

Merge preview

Version 0.9.4
This software is heavily under development, stay tuned.
Tested only with Firefox >= 3.5 and Google Chrome >=5

Role of filmographic data in EFG portal

Example: Vampyr, DE FR 1938

- Portal is not a common filmography but a search engine for digital material held in Europe´s film archives
- Filmographic data used to contextualise digital objects
- Shows only limited set of filmographic data held in EFG (names, titles which are directly related to digital objects)
- Initial portal concept: to exploit more extensive filmographic data for the functions that contextualise the digital material

RELATED NAMES

Dreyer, Carl Th. I Screenplay

Jul, Christen I Screenplay

Maté, Rudolph I Director of photography

Taldy, Tonka I Editor

Zeller, Wolfgang I Music composer

> **More**

RELATED TITLES

No matches

View on Danish National Filmography

Vampyr

Other title(s):	Vampyr (L'etrange aventure de David Gray)
	Allan Grays sælsomme hændelser
	Vampyr (Die seltsame Geschichte des Allan Gray)
Genre:	Feature film
Country:	DE
	FR
Year:	1932
Runtime:	75 min
Description:	En ung mand, Allan Gray, ankommer til en kro ude på landet. Samme aften får han besøg af en ukendt mand, der beder ham om at hjælpe med at frelse en ung pige, der er i fare. Hun viser sig senere at være den ukendte mands datter. Kort efter bliver den ukendte mand, der bor på et stort gods, myrdet, og den unge mand må selv prøve at komme den unge pige til undsætning. Derfor giver han blod til en transfusion, forhindrer den syge kvinde i at tage gift, befrier hendes søster, der er bundet til en seng, og drager til sidst bort med hende. Desuden er det lykkedes ham at standse den forbandelse, der hvilede over egnen, ved sammen med tjeneren at drive en jernpæl gennem en gammel kvindelig vampyrs hjerte. Derefter sørger tjeneren for, at hendes medskyldige, lægen, bliver kvalt - i mel (fra "Carl Th. Dreyer født Nilsson" af Martin Drouzy, 1982).
	A young man, Allan Gray, arrives at a countryside inn. At the very same evening an unknown man pays him a visit and asks him for help to save a young girl who is in great danger. As it turns out the girl is the unknown man's sister. Shortly after the visit the unknown man - resident at a great manor - is murdered, and Allan Gray has to come to the young girls rescue all alone. He gives blood to a transfusion, prevents the sick woman from taking poison, sets her sister who is tied to a bed free, and finally sets out with her. Furthermore he has succeeded in putting an end to the curse that has been upon the locality; together with the servant he has driven an iron pile into the heart of an old female vampire. The servant makes sure that her accomplice, the doctor, suffocates - in flour.
Provider:	Det Danske Filminstitut
Production company:	Film-Production Carl Dreyer
Director:	Dreyer, Carl Th.
Colour:	Black & White
Sound:	Without sound

Achievements

- Set up and tested workflows how authority records can be created in a common database with the help of semantic technology tools developed by EFG

- Increased quality of person and film work records in archives´ **local databases** (partners use the AFM tool to clean their doublets)

- Increased quality of name and title records in the **EFG database** (normalisation, cleaning and enrichment, represented data in unique format)

- Laid groundwork for a common filmography

- Created an understanding what data quality means in the scope of authority file building and maintenance within film archival community and beyond

Lessons learnt

- Maintaince of common registry of filmographic authority files cannot be assured by EFG (project expires in Aug 11) → need to follow-up on this task in other initiatives

- Projects like EFG can only come up with experiences and recommendations how authority files can be created and managed in the film archival domain

- Match & merge heterogneous data from 16 different data sources was not first priority in this pilot project → AFM tool to be further refined

- Further work needs to be done to create unique references for the names and titles managed by film archives and by EFG

EFG Content Checker

![EFG - european film gateway]

After EFG: Towards a common filmography?

"The common filmography will serve as an authority file of European film works, persons and corporate bodies, providing the highest possible precision in information and content retrieval". (Quote from EFG Description of Work)

Common registry of film works with persistent IDs is still task for film archives

➤ Could be built upon filmographic data aggregated in EFG

➤ More standardisation (FIAF Cataloguing Commission - rules, CEN Technical Committee - structure)

Should be matched and linked with authority files from national libraries in bilateral projects (e.g. GND in Germany)

➤ Film archives have to manage their persons in authority files

➤ Integrate unique IDs of library catalogues into film archival catalouges

Can be integrated in and linked with authority files from national libraries (e.g. GND in Germany).

➤ Should be postponed

➤ Requires more film-historic research

➤ Film archives would have to create authority files for corporations

Linked Open Data and Co-operations with Libraries

A lot of hard work ahead...

Himmelskibet (A trip to Mars), DK 1918

Thank you!