EFG´s data cleaning and enrichment tools

Europeana v1.0 WP3 meeting
Vienna, 29 March 2011

Francesca Schulze
Deutsches Filminstitut -DIF
f.schulze@deutsches-filminstitut.de

Project aims

- digital „showcase" for collections of Europe´s film archives and cinemathéques

  - Films and film clips, trailer (mostly: non-fiction)

  - Images: stills, posters, photos, set designs, drawings

  - Text material: correspondences, reviews, censorship documents

- make collections accessible through a common access point

- link Europe's film institutions with *Europeana*

- 3 years project (ends in August 2011)

- portal launch expected in May, URL: www.europeanfilmgateway.eu

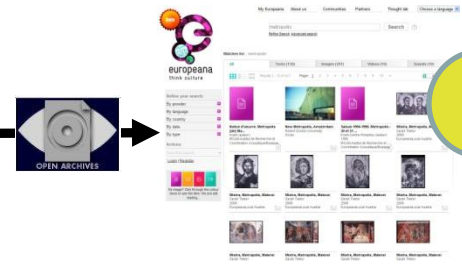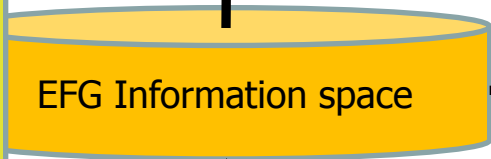- In August 2011: ca. 600.000 Items in *Europeana*

EFG
european film gateway

Infrastructure setup

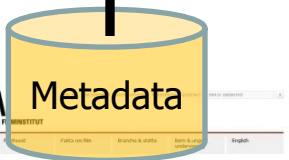Aggregation of filmographic, bibliographic and data on archival resources

2nd idea of EFG: common registry of film titles and persons active in the film domain:
120.000 film works
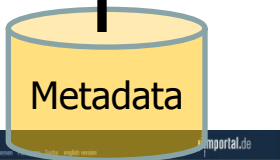230.000 persons
25.000 corporations
+ event records

Access to information about digital objects, related names and titles
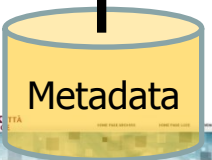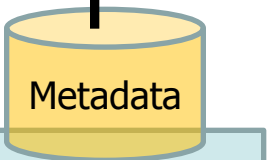(11 out of 16 Providers, 375.000 Items)

EFG Information space

340.000 Items in Europeana

Metadata

Metadata

Metadata

Metadata

Ca. 13 more sites
(16 Providers)

Danish Filmography (DFI)

Filmportal.de (DIF)

Luce site

Challenges

- harmonizing heterogenous and multilingual data from 16 film archives

- different quality of metadata

- no common standards for metadata schemas and vocabularies in film archival domain

- information about the same person or film work can come from different data sources due to numerous European co-productions

➢ complex metadata schema based on FRBR oriented Cinematographic Works Standard

  – www.filmstandards.org

Authority File Manager

- ➢ Data cleaning
- ➢ Tool to identify doublet records of the same film work or person (active in the film domain) in EFG Information Space

- • allows partners also to detect doublets within their own databases
- • merge doublet records to authority records
- • uses basic concepts of „Match & Merge" 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

Metadata Editor

➢ Data enrichment

➢ Tool allows partners to improve the quality of their metadata records

• encouraged archives to enrich their data locally with EFG relevant information

• most important enrichments: descriptions for non-audiovisual digital objects (also in English), establish relations to persons depicted on an image, creator, et. al.

• individual cataloguing plans

Metadata Editor

Vocabulary Checker

➢ Data cleaning
➢ Tool used to verify vocabulary matchings

- matching intellectual work by partner archives
- 46 EFG mini-vocabularies (value lists for attributes & semantic relationship types)
- FIAF glossary of filmographic terms, ISO country and language codes, Marc relator codes, et al.
- translated into 13 European languages
- available on EFG project website: www.europeanfilmgateway.eu/guidelines_and_standards.php

Vocabulary Checker

Open workshop on data quality and
semantic interoperability issues
in European film archives

30 May in Frankfurt

Thank you!
f.schulze@deutsches-filminstitut.de
www.europeanfilmgateway.eu