Data Interoperability and Curation: the European Film Gateway Experience

Michele Artini, Alessia Bardi, Federico Biagini, Franca Debole, Sandro La Bruzzo,
Paolo Manghi, Marko Mikulicic, Pasquale Savino, Franco Zoppi

Consiglio Nazionale delle Ricerche
Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
Via Moruzzi 1, 56124 Pisa, Italy

name.surname@isti.cnr.it

Abstract. Film archives, containing collections of cinema-related digital material, have been created in many European countries. Today, the EC Best Practice Network Project EFG (European Film Gateway) provides a single access point to 59 collections from 19 archives and across 14 European countries, for a total of 640,000 digital objects. This paper illustrates challenges and solutions in the realization of the EFG data infrastructure. These mainly concerned the curation and interoperability issues derived by the need of aggregating metadata from heterogeneous archives (different data models, hence metadata schemas, and exchange formats). EFG designed a common data model for movie information, onto which archives data models can be optimally mapped. It realizes a data infrastructure based on the D-NET software toolkit, capable of dealing with data collection, mapping, cleaning, indexing, and access provision through web portals or standard access protocols. To achieve its objectives EFG has extended D-NET with advanced tools for data curation.

Keywords: Data Infrastructure, Aggregation System, Metadata Formats, Data Interoperability, Data Curation, Data Cleansing, Audio Video, D-NET.

1 Introduction

Nowadays, many digital film archives are available in Europe, thanks to a significant effort performed in digitizing existing collections of images, videos, and cinema-related material (e.g., audio documents, photographs, posters, drawings, text documents). These archives make their collections available to the community through repository platforms or similar technologies, which support web portals to search, browse, and visualize cinema-related metadata and relative digital objects. Although the information dissemination service they offer is extremely useful, their autonomy still represents a limit to the urgent demand of immediate and global access to information required by today's communities.

adfa, p. 1, 2011.

The EFG (European Film Gateway) Best Practice Network [1], funded by the European Commission under the eContent*plus* programme [2], provides community users with a single entry point from which content of several archives can be searched in a uniform fashion, abstracting over their differences and peculiarities. Specifically, EFG delivers a data infrastructure whose aim is to aggregate content from the most prominent European film archives and cinematheques in order to make it available to end users and authorized third-party consumers, including Europeana (European Digital Library) [3]. The project started in September 2008 and was completed in October 2011 and its two-year continuation will kick-off on February 2012. It includes 20 partner institutions from 14 European countries, and today provides direct access to about 640,000 digital objects including films, photos, posters, drawings, and text documents, plus authority files for film works, persons and corporate bodies.

Although film archives contain similar digital movie-related objects, their data models (and relative metadata schemas) may be very different in structure and semantics, as well as their content be subject to errors or be duplicated. In this paper, we describe the solutions to the *data interoperability* and *data curation* challenges faced in EFG in order to deliver a unified, homogeneous, high-quality, and unambiguous European information space of movie metadata.

Data interoperability The EFG infrastructure has adopted a bottom-up approach to data aggregation where interoperability is achieved by (i) defining a common data model and relative metadata schema together with domain specific vocabularies, and (ii) implementing the technology to collect, transform onto the common schema, and harmonize metadata records collected from the archives. The EFG data infrastructure technology is powered by the D-NET [15] software toolkit, which provides a rich and customizable set of data management services capable of coping with issues such as metadata collection, storage, indexing, transformation, and cleaning. D-NET also offers services for the deployment of portals that can be configured according to the target community requirements, hence enabling end-users to search/browse the information space. Moreover, the D-NET toolkit includes mediation services for systems to access the space through standard protocols, such as OAI-PMH [22] and SRW/CQL, and several exchange formats.

Data curation Once metadata records are aggregated into a structurally and semantically homogenous information space, the EFG infrastructure enables archive experts to perform data curation actions by delivering easy-to-use tools for metadata validation, editing, de-duplication (e.g. the same persons and movies entities collected from different repositories). To this aim, the authors extended D-NET with services implementing the data curation functionalities for content and vocabulary checking, metadata editing, and authority file management (i.e., record de-duplication)

Paper outline: Section 2 gives an overview of the problem and introduces the adopted solution. Section 3 describes the main characteristics of the EFG common metadata schema. Section 4 describes the D-NET software toolkit. Section 5 presents the EFG D-NET-based infrastructure and its extension with D-NET data curation services. Finally, Section 6 concludes the paper.

2 Overview of the Problem and Adopted Solution

The EFG data infrastructure delivers two main requirements as identified by the user community:

- *Single access point to the European movie archives*: it supports advanced search and browse over all different types of collections (videos, images, textual documents), visualization of detailed metadata descriptions, and metadata export to third-party services, including Europeana.
- *High-quality metadata descriptions*: the EFG information space does not contain documents with poor descriptions and avoids duplication of information.

As mentioned in the introduction, these requisites are hindered by the highly heterogeneous nature of the archives. In fact, content of different archives generally conforms to different metadata models and XML schemas, whose structure may vary from complex element trees to simple flat sets of elements. Moreover, such content may describe different entities or the same entities, but with distinct semantics; e.g., different vocabularies of terms and format representation standards for dates, names, time durations.

To tackle such heterogeneity, EFG delivered two main outcomes: the EFG common data model and relative XML schema, onto which archive metadata records can be mapped; the EFG data infrastructure, whose services offer functionality for (i) collecting XML records from the archives and transforming them onto records matching the common XML metadata schema, and (ii) curating the resulting records by identifying and fixing semantic errors and duplicates. The data infrastructure was realized by adopting the D-NET Software Toolkit [15] and extending it with D-NET services for data curation.

The data ingestion workflow (sketched in **Fig. 1.**) consists of four phases and requires an interaction between domain experts and infrastructure administrators, adequately supported by the infrastructure services. These actors are driven by a detailed methodology, whose aim is to enable a controlled data ingestion life-cycle which will incrementally lead to the publication in production of a high-quality information space. Such workflow consists of four phases:

Phase 1: metadata mapping definition. Domain experts from the archives analyze the metadata they provide to determine how such information may structurally and semantically map onto the EFG metadata schema. The relative structural and semantic mapping rules are handed over to infrastructure administrators, who encode them in the form of D-NET scripts.

Phase 2: metadata transformation and cleaning. Archive metadata records are collected via OAI-PMH or FTP protocols to be processed through the mapping scripts produced in phase 1 and generate corresponding EFG records. The resulting records are not immediately available for access, but stored in a “pre-production” information space, where the Phase 3 of the workflow can take place. As we shall see, the Phase 1 and Phase 2 may be fired several times to refine the mapping rules and achieve the best metadata quality.

Phase 3: metadata quality control and enrichment. Records in the pre-production Information Space can be validated and inspected to identify mapping errors, mistakes (e.g., typos), and duplicates. Specifically, the Content Checker Tool can be used to verify that structural mapping was properly performed, the Vocabulary Checker Tool notifies data providers about EFG records not yet complying with the common vocabularies, and the Authority File Manager (AFM) identifies possible record duplicates. This quality control process may lead to the redefinition of the mapping rules (Phase 1), the adjustments of the mapping scripts (Phase 2), or to a subsequent data enrichment process. The Metadata Editor Tool enables curators to edit EFG records, while the AFM can fire record merge actions and effectively remove the duplicates.

Phase 4: metadata publishing. EFG records which passed Phase 3 are moved to the production Information Space, where they become visible from the EFG portal and can also be exported to third-party providers, such as Europeana.

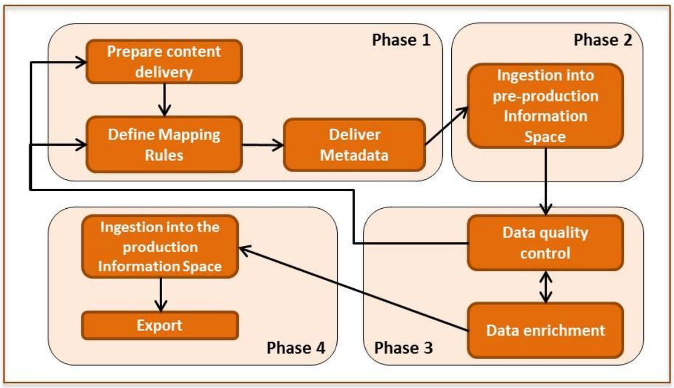


Fig. 1. Phases of the EFG data ingestion workflow

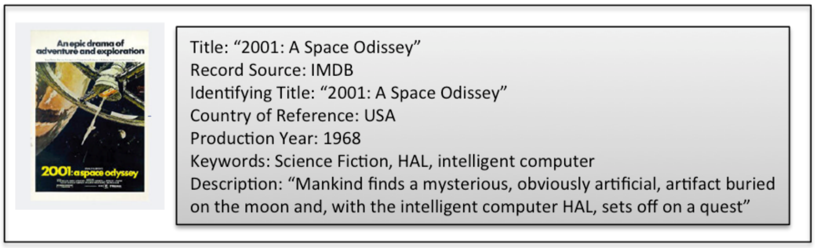
3 EFG Common Metadata Model and XML Schema

The EFG Common Metadata Model was designed after the analysis of the metadata models and schemas adopted within various organisations operating in the audio/video domain, starting from the data providers of the EFG consortium. This study took into consideration standards such as FRBR [4] and Dublin Core [5], as well as more film-specific standards such as the Cinematographic Works Standards EN 15907 [6]. As a result, eight interrelated entities have been defined in the EFG Common Metadata Model [19][24]:

- The *AVCreation* contains the properties of a cinematographic work: the film title, the record source (archive), the country of reference, the publication year, etc.

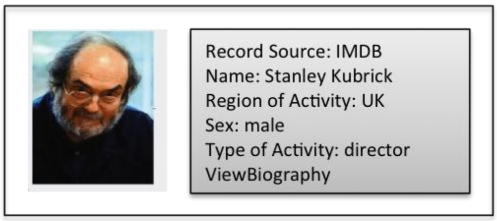
- The *AVManifestation* contains the information about the physical embodiment of an audiovisual creation. Examples are archival copies (analogue or digital) and database files. Properties of an AVManifestion include language, dimension, duration, coverage, format, rights holder, and provenance.
- The *NonAVCreation* describes all non audiovisual creations that can be represented in EFG. These are pictures, photos, correspondence, books or periodicals. The properties of NonAVCreations are: title, record source, keywords, description, date of creation and language.
- The *NonAVManifestation* entity keeps track of copies of non-audiovisual objects. It has properties such as type (e.g. text, image, sound), specific type (e.g. photograph, poster, letter), language, dates (i.e. a date or period associated with the issue of the manifestation), digital format (including its status, size, resolution), physical format, geographic scope, rights holder.
- The *Item* entity points to the digital file held in the source archive. Its attributes are isShownBy (i.e. the URL reference to the digital object on the content provider's web site), isShownAt (i.e. the URL reference of the object in its information context), digital format, provider and country.
- The *Agent* is defined as an entity that can perform an action. The model includes three agent types: Person, Corporate Body and Group. For example, the Person Agent has the following properties: name (composed of prefix, forename and family name), type of activity, date (which specifies the temporal properties of the person in relation with his activity), place (where the activity was performed), sex. Similar properties are defined for Corporate Body and Group.
- The *Event* is an entity that can occur within the lifecycle of an audiovisual or non-audiovisual creation. Examples of Events are Physical Event (e.g. a public screening or a broadcast), Decision Event (e.g. when a manifestation of a creation was evaluated by a censorship body), IPR registration, Award (i.e. the award obtained by an audiovisual creation or an agent), Production event (e.g. dates and places where castings took place, dates and locations of shooting).
- The *Collection* is defined as a compilation of creations (audiovisual or non-audiovisual).

In order to better illustrate the model and the relationships it defines among the above entities, we show a real-case example about the film "2001: A Space Odyssey" directed by Stanley Kubrik. We may have a record description of the AVCreation as follows:


An epic drama of adventure and exploration
2001: a space odyssey

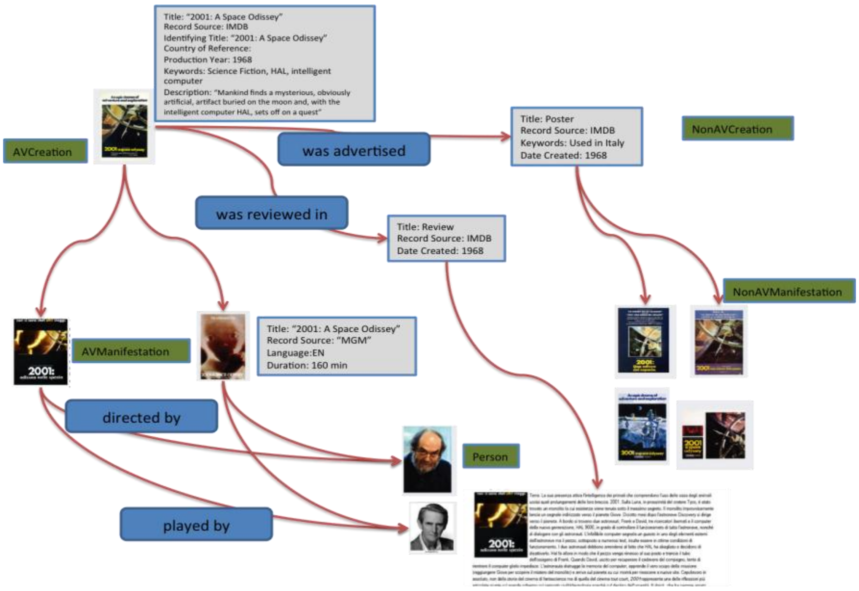
Title: "2001: A Space Odissey"
Record Source: IMDB
Identifying Title: "2001: A Space Odissey"
Country of Reference: USA
Production Year: 1968
Keywords: Science Fiction, HAL, intelligent computer
Description: "Mankind finds a mysterious, obviously artificial, artifact buried on the moon and, with the intelligent computer HAL, sets off on a quest"

The record description includes some metadata elements plus a thumbnail describing the AVCreation. We will have several AVManifestations associated to the AVCreation, such as all national versions of the movie, for example the Italian and the American versions. At the same time we may have several Agents related to this movie. As an example, we show a record description for the movie director, Stanley Kubrick:



Record Source: IMDB
Name: Stanley Kubrick
Region of Activity: UK
Sex: male
Type of Activity: director
ViewBiography

Furthermore we may have NonAVCreations such as posters and film reviews. All these entities are connected through relationships (see **Fig. 2**). The metadata record associated to each entity will be used to retrieve the archived object, while the relationships will be used to support browsing. As an example, it is possible to search for all movies directed by Stanley Kubrick in the ‘50s and browse all received awards, biographies of actors, etc.



Title: "2001: A Space Odissey"
Record Source: IMDB
Identifying Title: "2001: A Space Odissey"
Country of Reference:
Production Year: 1968
Keywords: Science Fiction, HAL, intelligent computer
Description: "Mankind finds a mysterious, obviously artificial, artifact buried on the moon and, with the intelligent computer HAL, sets off on a quest"

Title: Poster
Record Source: IMDB
Keywords: Used in Italy
Date Created: 1968

Title: Review
Record Source: IMDB
Date Created: 1968

Title: "2001: A Space Odissey"
Record Source: "MGM"
Language:EN
Duration: 160 min

Fig. 2. Example of metadata associated for the film "2001: A Space Odyssey"

The EFG Common Metadata XML Schema [19] implements the common model described so far. It defines XML element types and attributes for all the eight entities and their relevant properties. The common schema is conceived as the type union of eight XML schemas (one for each entity) in such a way that one EFG XML record represents one entity together with its relationships to other entities. Furthermore, the schema defines the so-called “controlled elements”, which are the XML elements whose values must comply with a given vocabulary of terms.

4 Enabling Data Infrastructures: the D-NET Software Toolkit

In the last decade, as witnessed by several national initiatives (e.g., BASE 7, DARE-net [8], OAIster [9]) and EC projects (e.g., Europeana [3], Bricks [10], ScholNet [14], DILIGENT [11], D4Science [12], DRIVER [16], OpenAIRE [17], CLARIN [13], HOPE [18]), the diffusion of Digital Libraries which took place in the last ten-twenty years in several communities, has been followed by an urgent need for integrating and aggregating content from such DLs to make it available through a single access point. In the last three Framework Programme calls, the European Union initiated the so called *knowledge infrastructure vision*, inspired by the same goal of unifying data resources of all kinds available in Europe. The idea was that of devising *data infrastructures*, which are environments through which several organizations can share, process, aggregate their data resources by adopting an economy of scale approach. Several technological solutions [20] were devised in such projects, to offer functionality for collecting data from heterogeneous data sources (e.g. repository systems, archives, databases), curating such data to form a homogeneous information space, and offering customized portal services to operate over such space; e.g. search, inference of references between publications, citation calculation, etc.

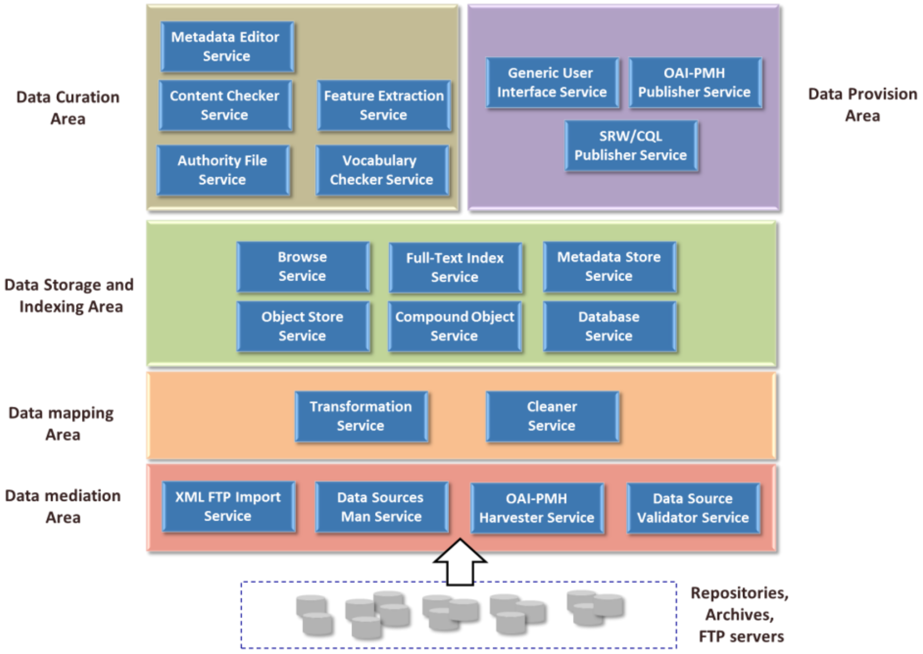


Fig. 3. D-NET service architecture

Of particular interest to Digital Libraries is the *D-NET software toolkit*, resulting from the experience of DRIVER, DRIVER-II, and OpenAIRE EC projects. D-NET is an open source solution specifically devised for the construction and operation of customized data infrastructures. D-NET provides a service-oriented framework where data infrastructures can be constructed in a LEGO-like approach, by selecting and properly combining the required D-NET services (such architectural concept was devised at CNR-ISTI by some of the authors of this paper). The resulting infrastructures are customizable (e.g., transformation into common metadata formats can be configured to match community preferences), extensible (e.g. new services can be integrated, to offer functionality not yet supported by D-NET), and scalable (e.g., storage and index replicas can be maintained and deployed on remote nodes to tackle multiple concurrent accesses or very-large data size). D-NET offers a rich set of services (see **Fig. 3**) targeting aspects such as data collection (mediation area), data mappings from formats to formats (mapping area), and data access (provision area). Services can be customized and combined to meet the data workflow requirements of a target user community. As proven by the several installations [15] and adoption in a number of European projects (DRIVER, DRIVER II, OpenAIRE, HOPE), D-NET represents an optimal and sustainable solution [21] for the realization of the EFG infrastructure. In the context of the EFG project, D-NET has been successfully extended with further generic and configurable services (curation area) for advanced curation and validation of XML metadata records.

5 EFG Data Infrastructure

The EFG data infrastructure consists of the D-NET services shown in **Fig. 3**, appropriately combined to support the data ingestion workflow presented in Section 2. In particular, the services in the Data Curation are resulted from the project activities. They were devised in order to meet the requirements of EFG archive partners, but engineered to support their functionalities when operating over arbitrary XML schemas.

5.1 Metadata Mapping Definition, Transformation, and Cleaning

Archives and their experts joining the EFG data infrastructure are supported with a methodology that facilitates the definition of *structural mappings* from their archive schema onto the EFG common metadata schema and *semantic mappings* from their vocabularies onto the common vocabularies. A mapping consists in a set of rules, which serve as input to the infrastructure administrators to configure the services in the Data Mapping Area. Here, the Transformator Service and the Cleaner Service run PERL scripts which parse, validate and transform the source records into EFG records according to the defined rules.

The Transformator Service is responsible for the application of *structural rules*. Such rules define the correspondence among elements and attributes of the archive schema and elements and attributes of the EFG schema. Structural mapping is not as trivial as it may seem, due to the fact that input XML records are typically mapped onto several interrelated EFG records, representing different EFG data model entities. More in detail, a structural mapping rule consist of the following information:

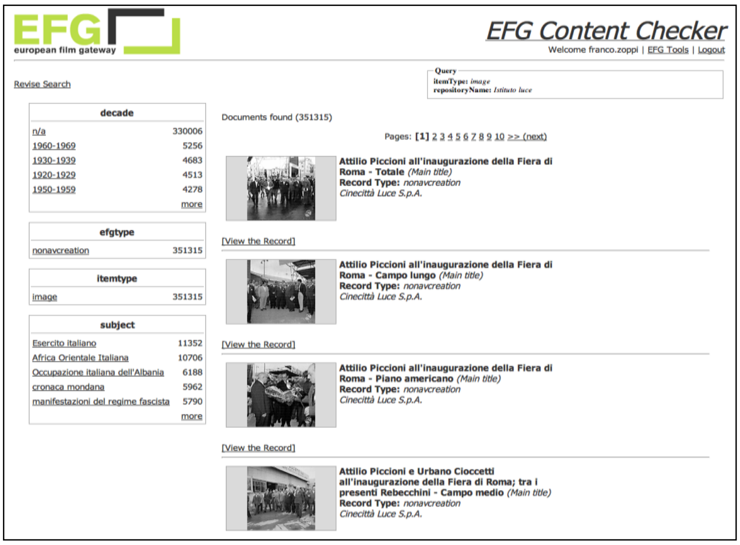
1. *Source element*: xpath identifying the schema element relative to the input value;
2. *Target element*: xpaths identifying the schema elements (and the sub-entity) onto which the source value should be mapped;
3. *Mandatory element*: states if the source element is mandatory (if not, the record is rejected);
4. *Element multiplicity*: states if the source element is repeatable;
5. *Comment*: description of the mapping rule.

The Cleaner Service is instead responsible for the application of *semantic rules*. Such rules identify an element of the archive schema and the corresponding element of the EFG schema (i.e., source element and target element of structural rules), and define the correspondence between the terms of the respective vocabularies.

5.2 Metadata quality control and enrichment

For the realization of the EFG data infrastructure the D-NET software toolkit has been extended with the following services, constituting the D-NET Data Curation Area.

Content Checker. The Content Checker (see **Fig. 4**) is a validation tool that allows low-level searching and browsing the pre-production Information Space in order to check if metadata records have been correctly harvested and mapped.



EFG european film gateway

EFG Content Checker

Welcome franco.zoppi | EFG Tools | Logout

Revise Search

Query
itemType: *image*
repositoryName: *Istituto luce*

decade	
n/a	330006
1960-1969	5256
1930-1939	4683
1920-1929	4513
1950-1959	4278
	more

efgtype	
nonavcreation	351315

itemtype	
image	351315

subject	
Esercito italiano	11352
Africa Orientale Italiana	10706
Occupazione italiana dell'Albania	6188
cronaca mondana	5962
manifestazioni del regime fascista	5790
	more

Documents found (351315)

Pages: **[1]** 2 3 4 5 6 7 8 9 10 >> (next)

Attilio Piccioni all'inaugurazione della Fiera di Roma - Totale *(Main title)*
Record Type: *nonavcreation*
Cinecittà Luce S.p.A.

[View the Record]

Attilio Piccioni all'inaugurazione della Fiera di Roma - Campo lungo *(Main title)*
Record Type: *nonavcreation*
Cinecittà Luce S.p.A.

[View the Record]

Attilio Piccioni all'inaugurazione della Fiera di Roma - Piano americano *(Main title)*
Record Type: *nonavcreation*
Cinecittà Luce S.p.A.

[View the Record]

Attilio Piccioni e Urbano Cioccetti all'inaugurazione della Fiera di Roma; tra i presenti Rebecchini - Campo medio *(Main title)*
Record Type: *nonavcreation*
Cinecittà Luce S.p.A.

Fig. 4. EFG content checker

Vocabulary Checker. The Vocabulary Checker gives access to the metadata records that do not satisfy the constraints imposed by the common metadata schema and vocabularies after the transformation and cleaning phases. The Vocabulary Checker displays the number, the types and the positions of errors in the records of the Information Space. Thanks to the browse by error typology functionality, curators can decide if an error can be solved directly in the Information Space via the Metadata Editor Tool or in the original source archive.

Metadata Editor Tool. The Metadata Editor Tool (MET) is a cataloguing tool for the enrichment of the Information Space. It allows data curators to add, edit and delete metadata records in the Information Space, as well as to establish relationships between existing (authority) records, even if coming from different sources. The MET is aware of controlled vocabularies, hence supports data curators while editing controlled elements by proposing a drop down list with all and only the terms defined by the associated controlled vocabulary. For example, let us suppose the Det Danske Filminstitut (DFI) EFG data provider provides a metadata record relative to the movie "*Olsen Banden over alle bjerge*", which features the actor *Ove Sprogøe*, but the actor is not mentioned in the metadata record. In order to make the record retrievable through the EFG portal to end users searching for "Ove Sprogøe", the movie record must be enriched with such information. The MET allows data curators to construct a relationship between the DFI movie metadata record and the person record, be the latter provided by harvesting other archives or created by data curators themselves.

Authority File Manager. The Authority File Manager (PACE [23]) is an advanced tool that curators can use to merge duplicate records and disambiguate the information space. The tool is capable of automatically identifying the pairs of records candidate for merging based on a multi-sort version of the sorted neighbourhood algorithm and a record similarity function that is customizable by data curators (they can chose between a range of similarity functions and assign different weights to the record fields). After one run of the candidate identification process, record pairs are displayed in descending order with respect to a 0…1 similarity distance. The curator has the responsibility of merging the two records (i.e., deciding if the two records are indeed representing the same entity). In the EFG scenario, the AFM has been configured to merge metadata records relative to persons and film works (AVCreation).

5.3 Metadata Publishing

The EFG Portal is available at [1]. Facilities like advanced metadata search and browse (by collection, provider, date, language and media type), search results filtering, video streaming, photo gallery and news highlights enhance the user experience in the phases of search and access. Moreover, D-NET offers services to export metadata records through OAI-PMH, OAI-ORE, and SRW/CQL protocols. EFG operate such services to automatically serve its information space to third-party consumers, above all the Europeana project [3], of which EFG is a direct feeder.

6 Conclusions and Future Work

We described the solutions adopted in the EFG Best Practice Network to achieve a complete integration of different national audio/video archives. The solution is based on the creation of a metadata schema that has, at the same time, the power to preserve the input metadata quality and the simplicity to enable simple mappings from all different archives. Metadata aggregation is based on the use of the D-NET software toolkit, a data infrastructure enabling software. D-NET offers services for metadata collection, transformation, and provision and its service-oriented framework allows for the addition of new services, to add domain specific missing functionalities. In EFG this resulted in the realization and integration of advanced curation and validation services: the Content Checker, the Vocabulary Checker, the Metadata Editor Tool and the Authority File Manager.

The current limitations of the EFG data infrastructure relate to the manual effort required in the phases of mapping rule definition and implementation and of metadata quality control and enrichment. Whilst some of the operations cannot be fully automatized, because archive administrators want to have control on data manipulation processes, we foresee some enhancements to (i) facilitate domain experts in the definition of mappings, (ii) (partly) automate the script-implementation of those mappings, and (iii) support experts and system administrator to ensure better metadata quality. Data provider experts currently define mappings by filling prefabricated Excel worksheets. Such files are then manually processed by infrastructure administrators to generate the corresponding transformation scripts. We could simplify this workflow by supporting data providers with a mapping definition tool, equipped with a GUI that shows a visual representation of their metadata schema and the common schema, and allows them to draw mappings by “dragging and dropping” elements of the first to elements of the second. The same tool could “generate” transformation scripts, at least when mappings can be reduced to a sequence of rule templates. Finally, we believe that the number of iterations of the transformation/cleaning and validation workflow could be reduced by providing a mapping test environment, where domain experts and infrastructure admins can verify the result of their mappings over a set of sample records.

7 Acknowledgements

This work is partly funded by the EFG Best Practice Networks project: grant agreement ECP 517006-EFG, call: FP7 EU eContentplus 2007. Its completion would have not been possible without the precious cooperation of Marco Rendina (Istituto Luce, Italy), Georg Eckes and Francesca Schultze (Deutsches Filminstitut, Germany) for the design of the data model.

8 References

1. European Film Gateway project, http://www.europeanfilmgateway.eu
2. eContentPlus framework, http://ec.europa.eu/information_society/activities/econtentplus

/index_en.htm

3. Europeana, http://www.europeana.eu
4. Functional requirements for bibliographic records: final report / IFLA Study Group on the Functional Requirements for Bibliographic Records. Saur, K.G. München : (UBCIM publications ; new series, vol. 19), 1998. ISBN 978-3-598-11382.
5. Metadata: The Foundations of Resource Description. Weibel, Stuart L. s.l. : D-Lib Magazine, 1995. Available online at: http://www.dlib.org/dlib/July95/07weibel.html.
6. Cinematographic Works Standard. Committee, Technical. 2005 : s.n.
7. BASE: Bielefeld Academic Search Engine http://www.base-search.net
8. DAREnet: Digital Academic Repositories, http://www.darenet.nl/
9. OAIster Official Site, http://www.oaister.org
10. Bricks Project, http://www.brickscommunity.org/
11. DILIGENT Project, http://diligent.ercim.eu/
12. D4Science Project, http://www.d4science.eu/
13. CLARIN Project, http://www.clarin.eu/
14. ScholNet Project, ftp://ftp.cordis.europa.eu/pub/ist/docs/rn/scholnet.pdf
15. D-NET Software Toolkit, http://www.d-net.research-infrastructures.eu
16. DRIVER Project, http://www.driver-community.eu/
17. OpenAIRE Project, http://www.openaire.eu/
18. HOPE Project, http://www.peoplesheritage.eu
19. Balzer, D., Debole, F., Savino, P.: Common interoperability schema for archival resources and filmographic descriptions, Deliverable D2.2 EFG Project
20. Manghi, P., Mikulicic, M., Candela, L., Artini, M., & Bardi, A. (2010). General-Purpose Digital Library Content Laboratory Systems. Research and Advanced Technology for Digital Libraries, 14th European Conference, ECDL 2010 (pp. 14-21). Springer.
21. Manghi, P., Mikulicic, M., Candela, L., Castelli, D., & Pagano, P. (2010). Realizing and Maintaining Aggregative Digital Library Systems: D-NET Software Toolkit and OAIster System. D-Lib Magazine , 16 (3/4).
22. Carl Lagoze and Herbert Van de Sompel. The making of the open archives initiative protocol for metadata harvesting. Library Hi Tech, 21(2):118 – 128, 2003.
23. Paolo Manghi, Marko Mikulicic, and Claudio Atzori. PACE: A General-Purpose Tool for Authority Control, Book: Metadata and Semantic Research, Communications in Computer and Information Science, 2011, Volume 240, Part 1, 80-92, Springer Berlin Heidelberg, ISBN 978-3-642-24731-6
24. Pasquale Savino, Franca Debole, and Eckes, Georg. *Searching and browsing film archives. The European Film Gateway Approach*. 4th International Congress on Science and Technology on the Safeguard of Cultural Heritage in the Mediterranean Basin, 6-8 December 2009, Cairo, Egypt.