FP7-ICT-2007-3-231161

Deliverable 6.2.2

European Digital Library implementation guidelines for audiovisual archives

Georgia Angelaki (European Digital Library Foundation)

10/05/2010

Document administrative table

Document Identifier	PP_WP6_D6.2.2_EuropeanaAVProvGuideline_R0 Release 0
Filename	PP_WP6_D6.2.2_EuropeanaAVProvGuidelines_R0_v1.00.odt
Workpackage and Task(s)	WP6 Audiovisual Preservation Competence Centre WP6T2. Competence Centre Activities
Authors (company)	Georgia Angelaki (Europeana Foundation)
Contributors (company)	Julie Verleyen (Europeana Foundation), Antoine Isaac (Europeana Foundation), Georg Eckes (DIF), Martin Gordon (RBB), Rob Turnock (Royal Holloway University of London)
Internal Reviewers (company)	Johan Oomen (B&G) Werner Bailer (JRS)
Date	10/05/2010
Status	Release
Type	Deliverable
Deliverable Nature	R - Report
Dissemination Level	PU - Public
Planned Deliv. Date	28/02/2010
Actual Deliv. Date	10/05/2010

Abstract

This report is addressed to audiovisual content holders and describes the conditions and the process for making available audiovisual content to Europeana (www.europeana.eu).

DOCUMENT HISTORY

Version	Date	Reason of change	Status	Distribution
0.10	07.01.2010	Outline sent to reviewers	Outline	Confidential
0.20	18.03.2010	First Draft	First full draft	Confidential
0.30	18.03.2010	Second Draft including comments from AAA	Second draft	Confidential
0.40	19.03.2010	Third Draft including input from EFG	Third full draft	Confidential
0.50	08.04.2010	Third Draft including comments by reviewers	Third full draft	Confidential
0.60	12.04.2010	Draft including comments by EUscreen	Fourth full draft	Confidential
0.80	03.05.2010	Final draft proofread version	Final draft	Confidential
1.00	10.05.2010	Final report	Publishable report	Public

Scope

The current deliverable is a public deliverable of the PrestoPRIME project. The project is developing a Competence Centre the goal of which is to act as a major information hub for practitioners, archivists and technology providers in the audiovisual (aka AV) domain with regards to digital preservation, digitisation and online accessibility to AV content.

More precisely, D6.2.2 is entitled “Guidelines for AV content providers to make their contents available to Europeana”. The aim of the deliverable is to describe the conditions and the process for making available audiovisual content to Europeana (www.europeana.eu). The deliverable is a product of Work Package 6 that is responsible for the setup of the Competence Centre and in particular of Task 2 that looks after the Competence Centre’s activities. The deliverable is developed by the EDL[1] Foundation that is also partner to the project and aims at explaining the process and the conditions of the submission of content to Europeana with a view to attracting more content providers to it.

The deliverable also outlines Europeana’s partner strategy of collaboration with other content aggregators. For this reason, we also present the most important developments with regards to the three major AV aggregators Europeana currently collaborates with for receiving content (European Film Gateway for cinema-related content, EUscreen for broadcast TV and DISMARC for audio and music content). Their activities with regards to their partner strategy, handling of copyright, workflow for submission, processing and presentation of contents and language policy are also presented. These activities highlight the various approaches to this work that, when cooperating an aggregator to submit data to Europeana, a potential content provider needs to be aware of.

[1] At the time of writing of the deliverable the EDL (European Digital Library) Foundation that is partner to the project had its statutes changed and was renamed to Europeana Foundation.

Executive summary

D6.2.2 is entitled “Guidelines for AV content providers to make their contents available to Europeana and is a public deliverable of the PrestoPRIME project. The project is developing a Competence Centre the goal of which is to act as an information hub for practitioners, archivists and technology providers in the AV domain with regards to digital preservation, digitisation and online accessibility to AV content.

The present document has the following goals:

- to describe the Europeana guidelines for submitting audiovisual content available
- to present the Europeana partner strategy based on collaboration with existing European content aggregators. New content providers wishing to submit data to Europeana are likely to first join one of the partner aggregators
- to present the major European cultural heritage aggregators in the audiovisual sector which are most relevant for potential new content providers, as well as their policies with regards to new partners and contents
- to raise awareness of the various requirements of the various aggregators with which a new content provider may need to comply, and also the fact that content is likely to be processed twice: once as part of the aggregator’s platform and once during the Europeana harvesting.

The deliverable is developed by the project partner the EDL Foundation. The Foundation is an organisation which opens worldwide access to European culture. Significant professional organisations representing museums, libraries, archives and AV content holders are members of the Board of the Foundation including FIAT/IFTA[2], IASA[3], ACE[4], CENL[5], ICOM-Europe[6] and EURBICA[7].

Europeana Foundation governs Europeana (www.europeana.eu), a portal (currently in beta form) that offers integrated search to the digitised collections of museums, libraries, archives and AV holders across Europe. The portal at present allows search of approximately 6 million digital objects. The interface is offered in all the official European languages and there are various personalisation functionalities enabled as well. Since the inception of the project, Europeana has received significant financial and political support by the European Commission and its launch was warmly welcomed by users worldwide.

The content in Europeana is classified as text, image, video or sound. Europe's AV holdings (in archives or other formal collections) have been estimated at 50 million hours of audio, video and film, most of it in analogue formats[8]. However, among the 6 million searchable objects in Europeana at present only 114.000 items are AV content, consisting of either video (around 91.500) or sound items (around 22.500). The vast majority of the content is images (4 million objects) and the remainder is text. The need to increase the presence of AV content in Europeana was clearly identified and this was set as a target in the content strategy for the operational release of the portal (Rhine release) later this year (2010).

The first section of the current document describes the benefits for AV content providers in making their content visible via Europeana, with the intention of encouraging European AV content holders

[2] The International Federation of Television Archives: http://www.fiatifta.org/cont/index.aspx

[3] International Association of Sound and Audiovisual Archives: http://www.iasa-web.org/

[4] Association des Cinémathèques Européennes: http://www.acefilm.de/

[5] Conference of European National Librarians: http://www.cenl.org/

[6] European regional organization of the International Council of Museums: http://www.icom-europe.org/

[7] General Assembly of the European Regional Branch of ICA, the International Council on Archives: http://www.eurbica.org/

[8] www.digitalpreservationeurope.eu

to make their data available via Europeana. The major position presented here is that Europeana is not aggregating the content itself (i.e. the digital object) but only the metadata which the information that describes the object and a thumbnail wherever that is possible, to act as a teaser for the content. Europeana offers free access to this information but in order to view the document in greater detail, the user is redirected to the content provider's website where the particular terms and conditions of access apply. AV content providers give access to their metadata and in return gain unprecedented visibility to an international audience via Europeana.

The following sections describe Europeana's organisational, technical and legal requirements for receiving data. The organisational aspects involve the compliance of a provider's content with Europeana's content policy. These aspects include the process of establishing contact with Europeana and the various agreements embodying the terms of collaboration between Europeana and a content provider or aggregator. The technical requirements deal with the compliance of the data with Europeana's data model and the process of submission and validation of the data. There is a significant workload involved in the preparation, mapping, harmonisation and submission of data, for which content providers and aggregators should be prepared.

As previously implied, Europeana is not working alone in the area of aggregation of content. On the contrary, Europeana is collaborating with aggregators of content operating on a local, national or international level and who are often focusing upon a particular topic or domain. Europeana guides new content providers into joining one of the existing aggregators (national, regional or thematic) which undertake the process of harvesting, harmonisation and (often) enrichment of the data before submission to Europeana.

Finally, in the last section, we discuss developments with regards to the three main aggregators in the audiovisual sector with whom Europeana currently collaborates. Each aggregator offers specialised access to particular audiovisual sub-domains:

- European Film Gateway focuses on film and cinema heritage
- EUscreen focuses on history and content of broadcast TV programs
- EuropeanaConnect (aka EuropeanaConnect Audio Aggregation Platform, using the existing DISMARC infrastructure) focuses upon sound and music archives.

New content providers should be aware of the various requirements of the various projects with regards to content submission. The main areas we explore in the present document are the content and partner strategy of each project, the data format, the content ingestion process and workflow, the language policy, the characteristics of the portal and the relationship of the project with Europeana. We have identified these as main areas of work that are of interest for a potential content provider.

The above-mentioned initiatives are partly funded by the European Commission and are made possible in the context of European projects. As projects they all have a short-term lifespan (around 3 years). At the tend of this term, they must have a sustainability model for maintaining and possibly expanding their operations. DISMARC that has been operational for some time now but EFG and EUscreen are relative newcomers – launched in September 2008 and October 2009 respectively. They both plan to launch their individual portals towards the end of 2010. However, many areas we are investigating are still under development and a lot of the information provided in the current document is based on planned activities. Last but not least, all projects have committed to making data accessible on their own portals as well as on Europeana.

With regards to the Europeana guidelines for the submission of data, it needs to be mentioned that Europeana is a generic portal for accessing culture. The baseline for the set up of the first version of the portal was to achieve search-interoperability across museums, libraries, archives and audiovisual collections. For this purpose a low common denominator was adopted as a principle for submitting data, as is reflected in Europeana's current data model. In this respect, at present, Europeana tackles all content, either textual, image, video or sound in the same way, as is

reflected in the Europeana Semantic Elements (ESE) data specifications[9]. The ESE specifications define the interoperability and therefore functionality of the upcoming Europeana Rhine release in 2010. The present document gives some recommendations specifically targeted at AV content but, by and large, the majority of the guidelines we describe are similar for all the domains.

A subsequent, major, Europeana release will be launched in Spring 2011. For this release, there is currently a new data model in development which is aimed at capturing more domain-specific knowledge. As a result, Europeana will produce new guidelines for the submission of content. The current deliverable will, therefore, be updated at the end of January next year to include any new Europeana procedures and details of the collaborating AV projects' related developments.

The European Digital Library Foundation expects to collaborate with PrestoPRIME in order to raise awareness among content providers about the benefits of opening up access to the audiovisual heritage. The Competence Centre has the potential to become a main point of reference for information and networking with regards to preservation, digitisation and access to audiovisual heritage. The current deliverable is the first resource with regards to access that is part of the PrestoPRIME pool of information. However, the Competence Centre could leverage the wealth of access guidelines and recommendations that Europeana and the collaborating projects are producing in areas such as copyright, interoperability and multilingualism, in order to further raise awareness and improve access in the audiovisual sector.

[9] The current version of ESE specifications is v3.2.2: http://version1.europeana.eu/c/document_library/get_file?uuid=c56f82a4-8191-42fa-9379-4d5ff8c4ff75&groupId=10602

Table of Contents

1. Introduction…… 9
2. Benefits for content providers and aggregators……..... 10
3. Europeana Content Objectives and Partner Strategy……..... 11
3.1 Europeana Content Objectives…….. 11
3.2 The Europeana Partner Strategy…… 11
3.3 Analysis of existing audiovisual content in Europeana……... 13
4. Workflow for a new content provider……. 18
4.1 Routes to Content: helping a new organisation identify the best route to submit data…… 19
4.2 Partner Application Form……. 21
4.3 Data submission workflow…… 21
4.4 Main aggregation routes of audiovisual content into Europeana……. 24
5. Technical and operational requirements for data submission…….. 25
5.1 Some basic principles…….. 25
5.2 What data are submitted?..... 26
5.3 Europeana Semantic Elements…….. 27
5.4 Metadata Harmonisation and Normalisation Guidelines……. 29
5.5 Content validation…….. 29
6. Legal Requirements……. 30
6.1 Data Aggregator and Data Provider Agreements…… 30
7. European audiovisual projects and their specifications……. 30
7.1 The European Film Gateway…….. 31
7.1.1 Introduction…….. 31
7.1.3 Network strategy…… 31
7.1.4 Type of content……... 32
7.1.5 EFG Data format…… 33
7.1.6 Content ingestion workflow and tools…….. 35
7.1.7 EFG portal specifications…….. 36
7.1.8 EFG and Europeana…….. 36
7.2 EUscreen…… 37
7.2.1 Introduction…….. 37
7.2.2 Network and partner strategy……... 38
7.2.3 Types of content and content strategy……..... 39
7.2.4 EUscreen data format……..... 41
7.2.5 Content ingestion architecture, workflow and tools……... 41
7.2.6 Multilinguality and language resources used…… 42
7.2.7 Online accessibility of content – EUscreen front-end specifications…….. 43
7.2.8 EUscreen and Europeana…… 43
7.2.9 Handling of Intellectual Property Rights Issues…… 43
7. 2.10 Contacts……..... 44
7.3 DISMARC…… 44
7.3.1 Introduction…….. 44
7.3.2 Types of content…… 45
7.3.3 Network and partner strategy……... 45
7.3.4 AAP data format…… 45
7.3.5 Language resources used……..... 47
7.3.6 Online accessibility of content – AAP portal specifications…… 47
7.3.7 AAP and Europeana…….. 47
7.3.8 Other guidelines and tools (IPR, digitisation, normalisation and enrichment guidelines,

etc)…….. 48
7.3.9 Contacts……... 48

8. Europeana and the PrestoPRIME Competence Centre…… 48

9. Conclusions and future developments……..... 49
10. Annex……..... 50

1. Introduction

Europeana is Europe's multimedia online library, museum and archive, and currently grants access to the cultural holdings of EU's twenty-seven member states. It includes more than 6 million books, maps, recordings, photographs, archival documents, paintings and films from national libraries and cultural institutions.

Europeana's goal is to the opening up of new ways of exploring Europe's heritage. Europeana offers free access to collections and treasures via a single web portal, available in all official EU languages. Europeana is currently in development - the full service will launch later in 2010. The operational Europeana portal will provide improved search functionality and access to over 10 million objects. A follow-up release in 2011 will showcase multilingual and semantic web features.

Europeana was launched in November 2008 by the European President José Manuel Barroso at the Council of Ministers of Culture. The prototype featured around 2 million objects coming from 100 institutions representing around 1000 organisations in Europe. The response from the press and from the users was immediate and overwhelmingly positive as testimonials show: "*Sirs, ..Europeana.eu is a discovery of a new world of culture and the Arts. Europeana is the universalisation of the Arts that until now was the privilege of few Europeans. Congratulations*". The first online user survey was run between May 6 and 26 2009; overall satisfaction was very high. Out of the 3,204 respondents, 60% rated features and functions as "good" or "excellent".

However, the first analysis of the content was also alarming: 70% of content came from 4 providers, 82% of content comes from 4 countries and 77% of content was classified as 'image'. In the future, the major challenge for Europeana is how to engage all cultural heritage content providers across Europe and manage to harvest, index, harmonise, enrich and make this content available in a sustainable, robust and user-friendly way to users world-wide.

Content in Europeana is in four formats: text, image, video and sound. All content providers are asked to specify in their metadata to which category an object belongs. This is part of the Europeana qualifier fields that was added to enable facet browsing of content following the results of user studies that showed users' preference for this functionality. Video and sound are characterised by the time dimension and can range from extracts of films to fragments or complete TV programmes, political speeches and academic lectures, footage, radio programmes, concerts, sounds, etc. Europeana does not impose any restrictions over theme or genre as long as the content comes from a partner considered to be a trusted source.

Europeana acknowledges the importance of AV cultural heritage and the present document aims at helping minimise the infamous "*20th century black hole*". Audiovisual documents such as films, radio and television programmes, audio and video recordings contain the primary records of the 20th and 21st centuries. '*Transcending language and cultural boundaries, appealing immediately to the eye and the ear, to the literate and illiterate, audiovisual documents have transformed society by becoming a permanent complement to the traditional written record*' - this extract from Unesco's website on the World Day for AV Heritage[10] outlines the importance of this recent domain of cultural heritage for citizens worldwide.

Europe's AV holdings (in archives or other formal collections) have been estimated at 50 million hours of audio, video and film, most of it in analogue formats[11]. However, among the 6 million currently-searchable objects in Europeana[12], only 114.000 items are AV content (videos 91.500, sound items 22.500). The vast majority of content is images (4m), the remainder is text. Another

[10] World Day of Audiovisual Heritage is celebrated on the 27th of October: http://portal.unesco.org/ci/en/ev.php-URL_ID=25563&URL_DO=DO_TOPIC&URL_SECTION=201.html

[11] http://www.digitalpreservationeurope.eu/

[12] the figures reflect the status as of March 2010

downside is that AV content in Europeana comes from no more than 20 European providers and aggregators, and even the vast majority of this content (around 78.000 objects) comes from one provider: the Institut National de l'Audiovisuel. This is not a favourable picture and Europeana has clearly set as a target in the content strategy for the operational release of the portal (Rhine release) later this year (2010) to increase the AV content accessible.

The current document aims at providing guidelines targeted in particular to audiovisual content holders to help them make available their contents through Europeana.

2. Benefits for content providers and aggregators

The web has opened up many new opportunities for AV content holders to make available their contents for commercial or non-for profit use and re-use. Irrespective of the mission and the ways AV content providers exploit their assets online, Europeana offers a unique resource that drives traffic to their websites and adds value to their contents.

The biggest fear of content providers is that they will be forced to give away digital content. However, Europeana only harvests the metadata describing a digital object, a thumbnail or preview that represents it wherever possible and a link to the digital object's location online in the provider's or aggregator's system. A user is therefore directed to the provider's or aggregator's site for viewing the digital object in detail.

As a prestigious portal with a recognisable brand, Europeana gives unique visibility to a content provider's collections worldwide. Videoactive's[13] traffic statistics, for example, showed that in October 2009 60% of the total traffic to the site was coming from Europeana. This is a very important incentive for audiovisual content providers involved in both non-profit and commercial exploitation of their content. By harvesting and exposing metadata to search engines, Europeana enhances the routes users follow to reach a provider's content. Without this possibility, users would have to know exactly where to find the material they are looking for and visit individual sites.

Europeana adds value to the content by juxtaposing related images, texts, videos and audio items, therefore, 'repatriating' geographically dispersed content into a single, coherent and contextual virtual space. As a result, Europeana enriches users' experience giving them the opportunity to study dispersed, multi-format content in one place. The multilingual interface makes it possible for users to search in their native language and retrieve objects that are otherwise inaccessible to them.

Content providers will also be able to receive enriched Europeana data and to integrate the data in their local websites and portals through the search APIs, therefore adding new value to their own services. These APIs will be developed in the course of 2010.

Furthermore, Europeana has received significant political and financial support from the European Commission since the start. The importance attributed to Europeana at the European level is also a driving force for Europe-wide governments to fund digitisation projects and aggregation at the national level.

Knowledge transfer is a key reason for being part of the Europeana network as well. There are a number of critical issues that all European content providers and aggregators deal with, namely object modelling, semantic and technical interoperability of data, multilingual access, IPR and business models for sustainable access. Europeana works with digital library experts, thinkers and practitioners from Europe and the States in these areas and knowledge is shared across the network via workshops, publications, seminars and

[13] www.videoactive.eu is a predecessor project to EUscreen

conferences. CCPA, the Council of Content Providers and Aggregators is responsible for the diffusion of this knowledge across all cultural heritage institutions Europe-wide.

3. Europeana Content Objectives and Partner Strategy

3.1 Europeana Content Objectives

As mentioned earlier, the content included in the first version of Europeana came from a small amount of content providers, a 'coalition of the willing' committed to developing Europeana as a proof of concept. An analysis of the content was performed in the summer 2009 to inform the development of a strategy for the future. As a result, the Content Strategy[14], including the relevant Partner Strategy, was released in August 2009. It set out content objectives for Europeana as well as the organisational and procedural aspects involved in reaching these objectives.

The initial content analysis revealed that almost 50% of the content came from French institutions and that only a handful of countries had significant presence: Germany (16%), the Netherlands (8%), and the United Kingdom (8%). All other countries were represented with 5% or less each. With regards to formats, 77% of content is classified as an image, 20% is text, video makes up almost 2.5% and less then 0.5% of the content is audio. The content strategy aims at giving priority to countries with low representation and gives "special attention to audio and video collections in order to increase the number of items available in those formats".

The **Content Strategy** aims were therefore set as following:
• reach 10m objects in Europeana by 2010;
• leverage the projects related to Europeana that contribute content and technology;
• ensure a wide variety of content in Europeana across the domains and Member States.

3.2 The Europeana Partner Strategy

The main Partner Strategy goals are to ensure a wide support and participation from institutions and aggregators for Europeana, Member States and other stakeholders. The Europeana Partner Strategy, therefore introduces a model of aggregation which is central to the development and success of Europeana.

A content provider for Europeana can be any organisation that provides digital content for access via Europeana and/or the metadata that enables access. The amount and type of content, the technical infrastructure, output formats and available documentation can vary significantly among content providers.

It is, however, practically impossible for Europeana to work individually with every European content provider due to the enormous amount of work in the harmonisation and normalisation of the metadata needed. Europeana therefore works with an intermediate layer of content providers, namely the aggregators.

An Aggregator is an organization that collects metadata from its group of content providers and transmits them to Europeana, which helps content providers with guidance on conformance with Europeana norms and which converts metadata if necessary. The aggregator also supports the content providers with administration, operations and training.

Aggregators can be grouped in horizontal or cross-domain aggregators and vertical or

[14] The content strategy can be found here: http://version1.europeana.eu/c/document_library/get_file?uuid=b7b24d45-116e-442f-8b85-fbf931ebee72&groupId=10602

domain-specific aggregator:

• **Horizontal aggregators** aggregate content across domains, for instance national aggregators like Culture.fr or regional aggregators such as Erfgoedplus.be.
• **Vertical aggregators** aggregate content from a single domain**.** Examples of vertical aggregators are TheEuropeanLibrary.org and the European Film Gateway. Vertical aggregators can work at regional, national or international levels.

A particular type of aggregator to which the partner strategy awards special attention is the Europeana collaborating projects. The European Commission has co-funded a number of European projects – the Europeana Group[15] - that help the different heritage sectors tackle issues such as IPR, metadata standardisation, interoperability and IPR. The goal is to improve online access to the cultural and scientific heritage, and to integrate the aggregated content for delivery through individual thematic portals and Europeana. We will see three of these content aggregators in more detail later in this document: European Film Gateway, EUscreen and DISMARC.

Other than EFG, EUscreen and DISMARC that we will see further down in detail, Europeana Group of projects includes:

- Europeana Local (europeanalocal.eu), a project which does not create a portal itself but which helps local and regional museums, archives, audio visual collections and libraries to create local aggregators and bring their content to Europeana.
- ATHENA (www.athenaeurope.org) is an aggregator mainly targeting museums; does not create a portal but makes available the aggregated content through Europeana.
- APEnet (http://www.apenet.eu/) is initiated by twelve national European archives and will create a portal to make archival material more accessible.
- BHL Europe (http://www.bhl-europe.eu/) improves the interoperability of Europe´s biodiversity heritage libraries by introducing common standards and creates a portal to facilitate the search for taxonomy-specific and biodiversity information.
- Europeana Travel's (http://www.europeanatravel.eu/) overall objective is to digitise content from Europe´s national and research libraries on the theme of travel and tourism and to make this content available via Europeana and local libraries' websites.
- MIMO - Musical Instrument Museums Online (http://www.mimo-project.eu/) digitises content and creates a common access portal for musical instruments.
- JUDAICA - Jewish Urban Digital European Integrated Cultural Archive (http://judaica-europeana.eu/) identifies content in European Institutions demonstrating the Jewish contribution to the cities of Europe. Content will be digitised and accessible through Europeana.
- CARARE (http://www.carare.eu/) will develop an aggregation service for the archaeology and architectural heritage domain, and will establish point of access to 3D/VR objects.
- Heritage Of People's Europe (HOPE) is a partnership of European social history institutions aimed at improving access to digital content about the history of the labour movement.
- Europeana Regia is a collaborative initiative between European libraries for the digitisation of royal manuscripts in Medieval and Renaissance Europe and providing the access to this content via Europeana.
- ASSETS – Advanced Search Services and Technological Solutions for the European Digital Library will develop, implement and deploy large-scale services focusing on searching, browsing and interfaces

As mentioned previously, an aggregator may or may not have a portal where the content is made accessible to the public. If the aggregator's portal is not accessible for the public this is called a 'dark portal' or 'dark aggregator'. An aggregator can act as a repository, storing the digital items,

[15] See the Europeana Group Homepage for an introduction to all the projects: http://group.europeana.eu/web/guest;jsessionid=43A5A6D43FC236153B8997BC24490AB0

but it can also only collect metadata with a link to the digital item. Thematic aggregators can either be domain specific or cross-domain.

Aggregators, on a national, regional or vertical level, play a key role not only in aggregating content, but also in the organisational structure, standardisation of content, services to end-users and future sustainability of Europeana and related projects and aggregators. They have the additional possibility of contributing to the European knowledge base for the creation of discovery and search services to bring cultural and scientific heritage to all users. These roles and their importance and relevance to Europeana are outlined below.

Within the Europeana organisational model aggregators play a key role in various fields:

• Disseminating the vision and objectives of Europeana to their network of institutions to create wide support for Europeana and to engage as many institutions in Europeana as possible
• Providing valuable feedback about the issues and discussions in their field.
• Promoting and implementing of standards further along the content provision chain.
• Providing domain specific expertise and skills to institutions and Europeana.

Content providers and aggregators participate in the Europeana decision- making and in the setting out of the strategies and policies through their membership in the CCPA- the Council of Content Providers and Aggregators[16].

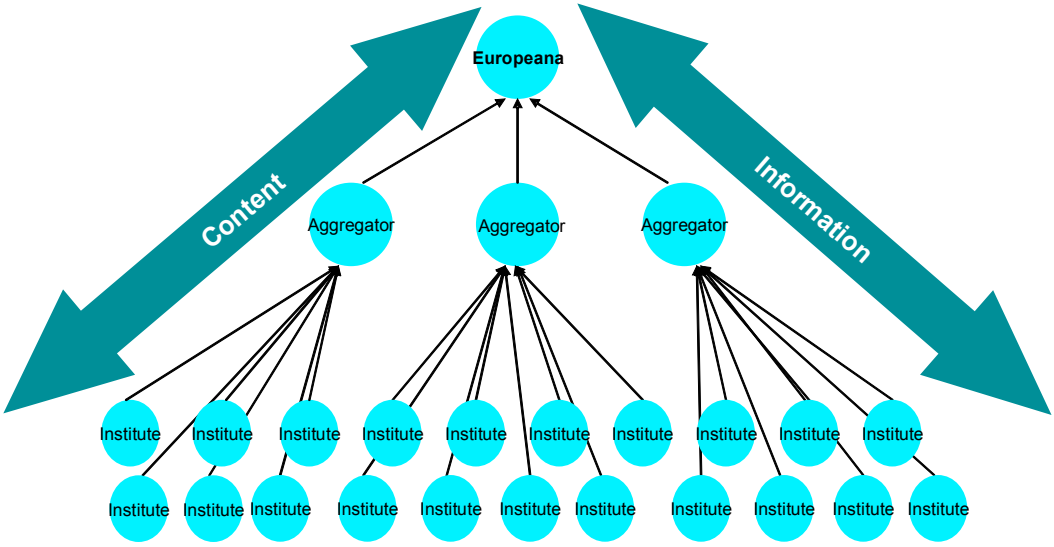

Figure 1: aggregators in the Europeana organisational model

3.3 Analysis of existing audiovisual content in Europeana

As of mid-February there were 91,796 videos in Europeana and 22,518 sound objects. These contents came mainly from the AV content providers that are members of the network of Europeana v1.0[17], the European Commission-funded project whose aim is to deliver the operational Europeana. We present the most important of them here, based on the size of the collections made available to Europeana.

[16] http://group.europeana.eu/web/guest/council/

[17] http://version1.europeana.eu/web/europeana-project/

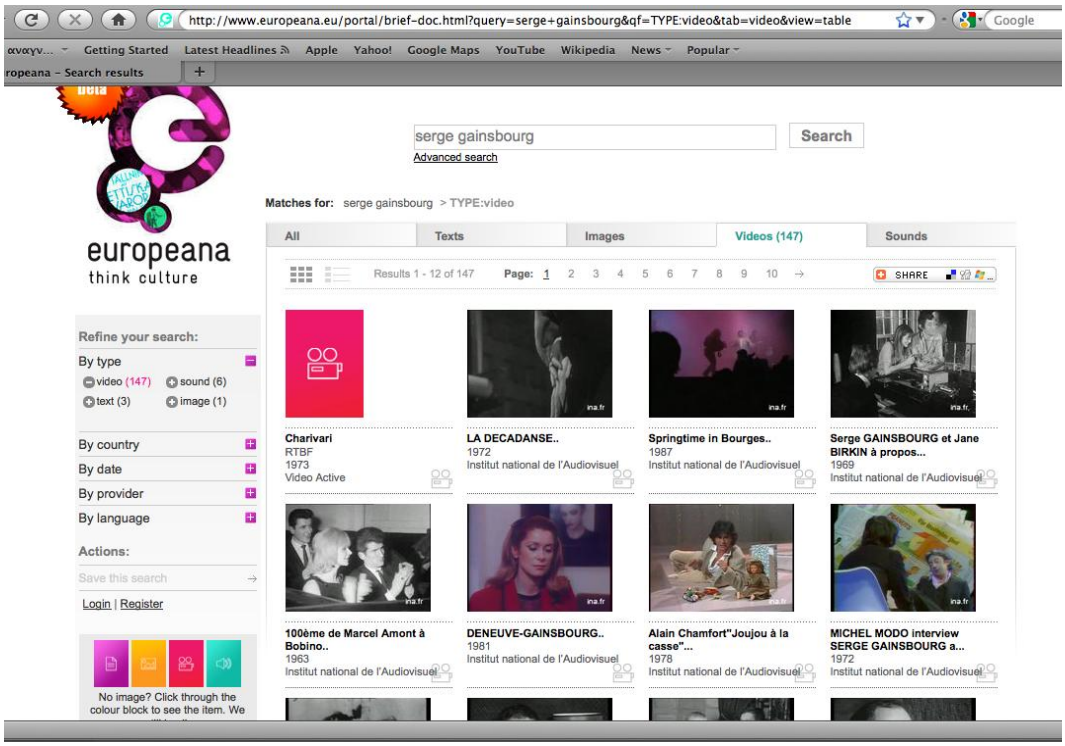

Figure 2: searching for video on Europeana

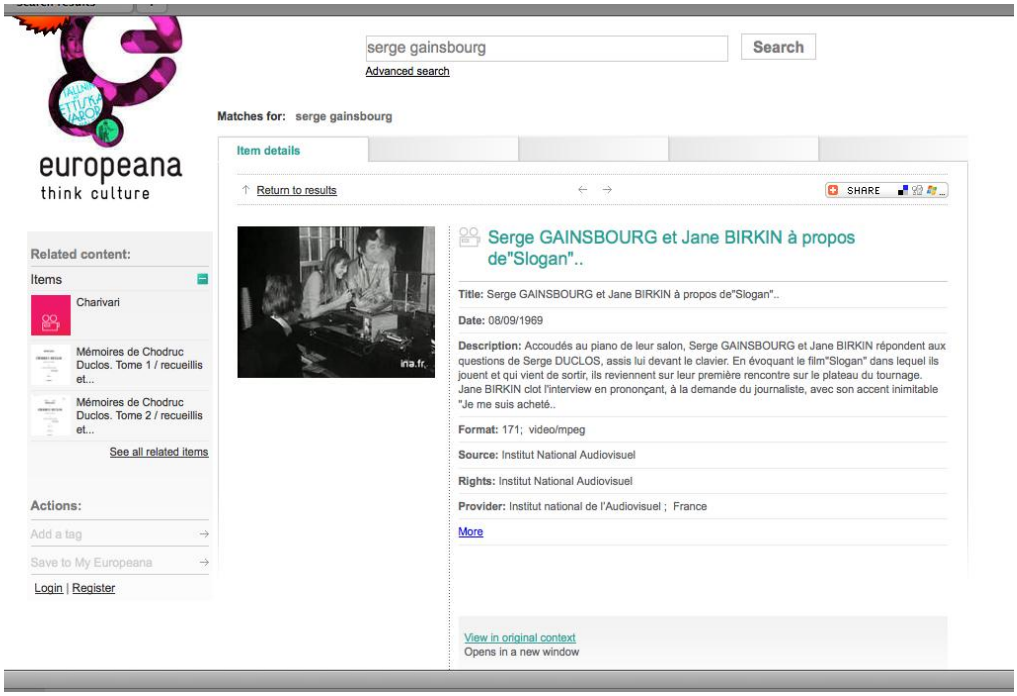

Figure 3: detailed results' page in Europeana

As of March 2010 the principal audiovisual content providers were:

Institut National de l'Audiovisuel (http://www.ina.fr/): 73.091 videos, 6.000 audio items

> INA is the national audiovisual archive of France and it's mission is to store and preserve the French Audiovisual heritage. Its contents consist of over three million hours of programs from the archiving of all TV (the last 60 years) and radio (70 years) broadcasting, as well as film newsreels between 1940 and 1969. It is also the legal deposit institution for such content and it also receives content from private companies and channels (i.e. National Geographic, TF1 …)

More than 25.000 hours[18] of TV and radio material are made available on INA's site (www.ina.fr) including the programs produced and transmitted by the public broadcasters (ORTF, Antenne 2, FR3, TF1 before privatisation, France 2, France 3, Radio France) and programs created by the production companies which have bestowed the rights of exploitation to INA. 80% of the archives are available free-of-charge, while for the remaining 20% only the first few minutes are free. All the programs apart from the special fonds are watermarked and can be downloaded or burned on DVD upon order from http://www.boutique.ina.fr/. The latter also sells some programs and DVDs compiled or created by INA. Any prices charged help remunerate rights holders and support INA's mission. As well as downloading, a number of other functionalities are available for users including the possibility of sharing, tagging, annotation and creation of playlists.

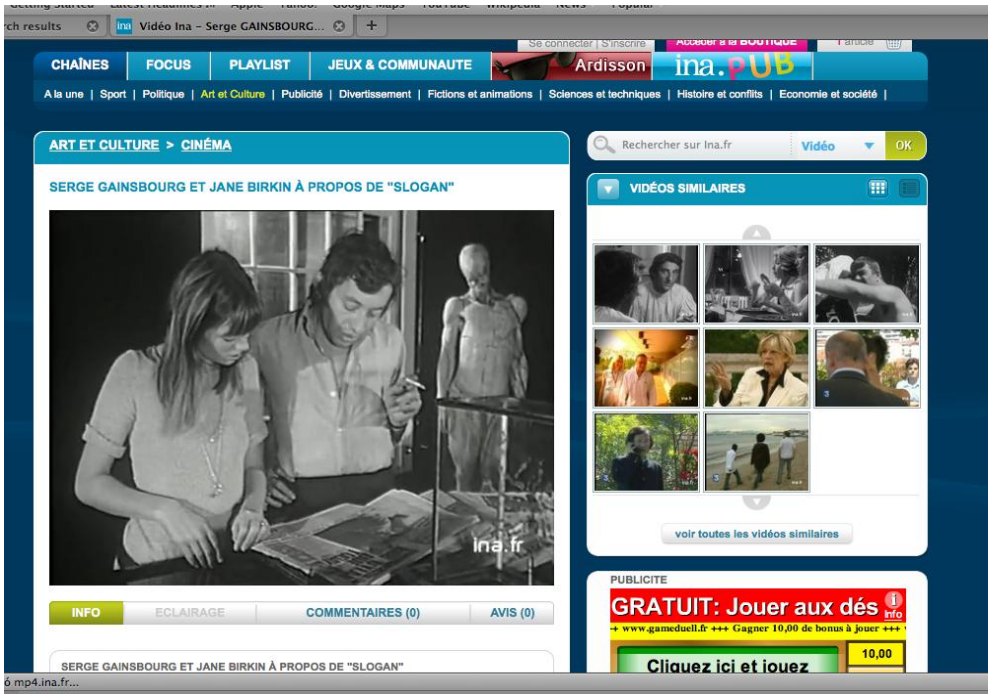

Figure 4: video display on the INA.fr website

The formats of the videos are
• MPEG4 SP (mp4)
• 320x240 pixels
380 kbits/s (340 kbits/s vidéo + 40 kbits/s audio aac mono)

and for audio:
•MPEG2 Layer 3 (mp3)
•24 kbits/s CBR
•Joint-stereo at 22050 Hz

The Adobe Flash player 9 (v 9.0.115.0 minimum) is required to view the videos

Scran (www.scran.ac.uk) : 6.292 videos, 226 audio items on Europeana
Scran is a local aggregator of digital collections in Scotland and England. It is part of the Royal Commission on the Ancient and Historical Monuments of Scotland and aims to provide educational access to digital materials representing material culture and history. It is one of the largest educational online services in the UK supporting over 4,000 schools, libraries, colleges and universities.

[18] http://boutique.ina.fr/statique/affiche/page/faq

The learning resource service hosts 360,000 images, movies and sounds from museums, galleries, archives and the media. It can be used as a superior form of clip art or for particular learning applications. Images of tigers, Charlie Chaplin, Sean Connery, a Degas, a Dali, images of war or whaling, standing stones, a pint of beer, an integrated circuit, or line drawings of an acorn or an adrenal gland demonstrate the range of images.

Materials on Scran are either copyright or have other rights associated with them, and subscription to the service is necessary to have access to full size images, videos or sound materials. Different rights are attributed to end-users, academic institutions being members of a network or other communities which is based on different licenses.

Scran is only making low resolution images available in Europeana. In order to view the full-object a user needs to buy a license or to become a member of an existing Scran-licenced network.

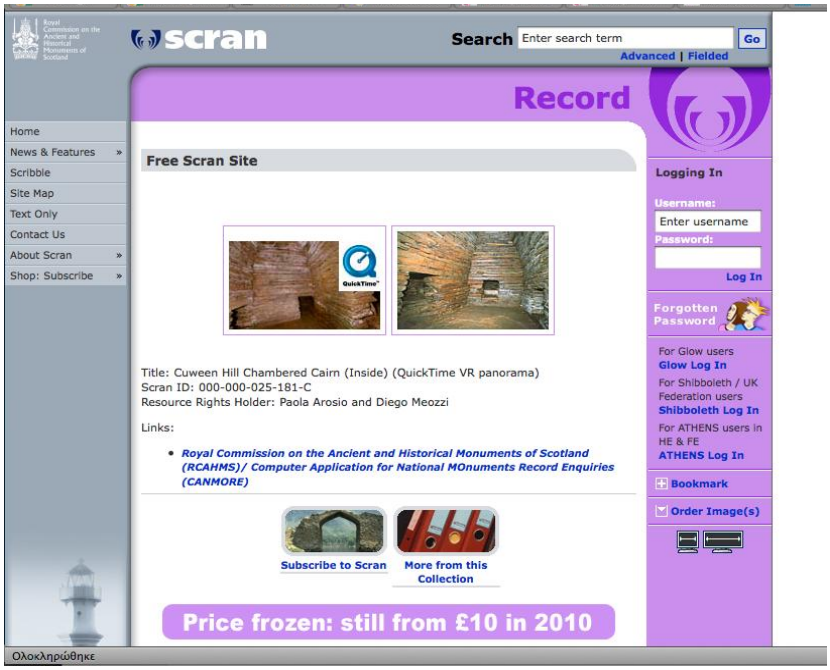

Figure 5: landing page on the SCRAN website

While respecting partners' copyright conditions, Europeana however, favours direct access at least to low resolution or watermarked content for private end-user viewing purposes online.

Centre Virtuel de la Connaissance sur l'Europe (http://www.cvce.lu/): 1890 videos, 811 audio items

The CVCE is a Luxembourg-based public undertaking supported by the Ministry of Higher Education and Research. The European NAvigator (ENA) digital library (www.ena.lu), is designed and developed by the Centre Virtuel de la Connaissance sur l'Europe (CVCE), and contains over 16.000 documents about the European integration process. Key post-war historical events and European organisations are presented chronologically and thematically, and are illustrated by a wide range of archive documents, as well as interviews conducted by the CVCE with people who have been actively involved in European integration.

In this digital library, pupils and students, teachers, researchers, and anyone interested in the European integration process can find original material such as photos, audio and video clips, press articles and cartoons, together with explanatory synopses, tables and

interactive maps and diagrams. The material included in ENA's vast and varied documentary resources is selected, created, processed and validated by a multidisciplinary team of specialists in European integration.

Search and consultation of the content is possible thanks to an online multimedia application: Multimedia Content Explorer (MCE) developed by CVCE. Audio and video files are streamed and optimised according to connection speed; all media are rendered in exactly the same way, regardless of the type of machine, operating system or browser used;

Online consultation is free of charge.

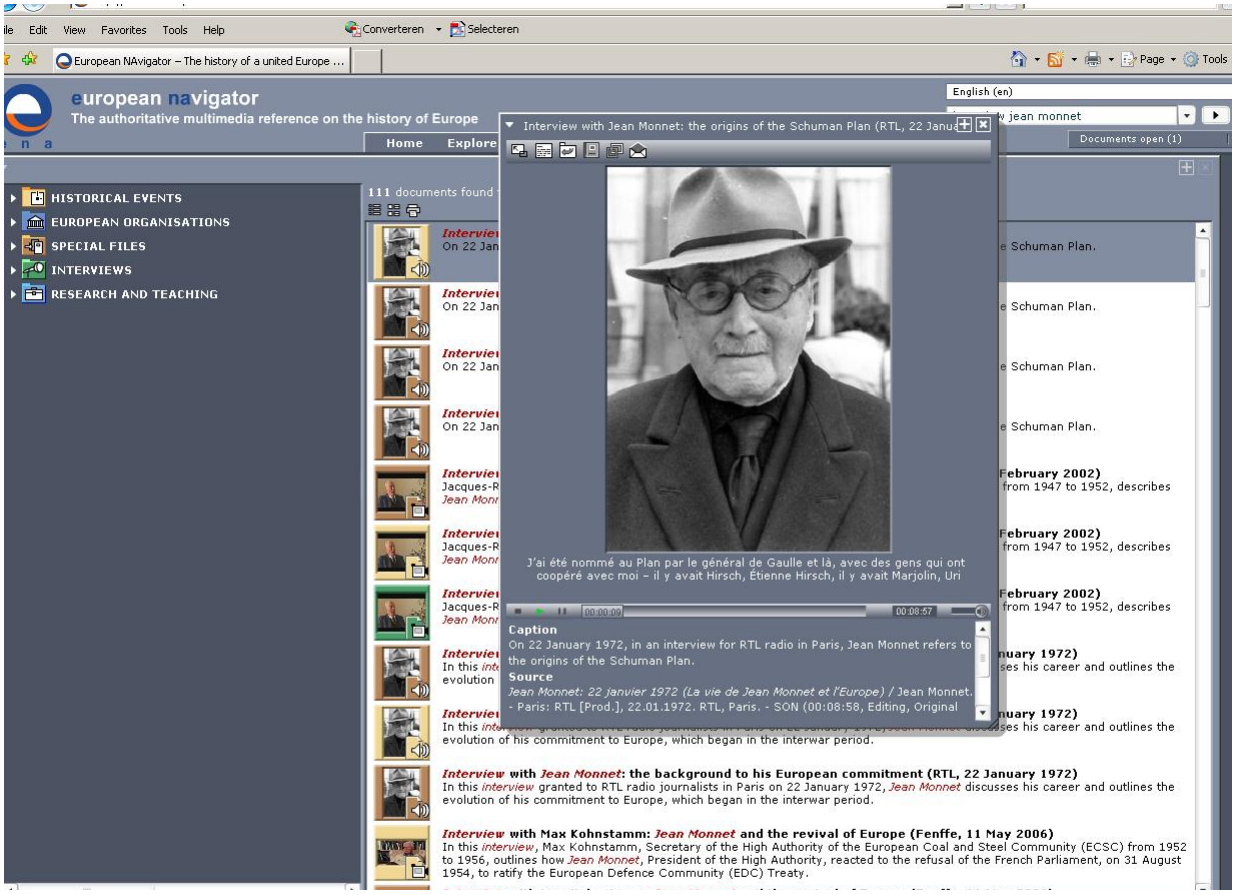

Figure 6: Audio rendering on CVCE

IRCAM-Institut de Recherche et Coordination Acoustique/Musique, France (http://www.ircam.fr/?L=1)

IRCAM makes the contents of the “Contemporary Music Portal” available to Europeana. The portal is curated by the Centre de documentation de la musique contemporaine (Cdmc) and aggregates contents of various collections from La Cité de la musique; Le Conservatoire national supérieur de musique et de danse de Paris; L'Ensemble intercontemporain; L'Institut de recherche et coordination acoustique/musique et La Médiathèque musicale Mahler. Metadata records from the providers have been harvested using the OAI-PMH protocol and metadata was made available as MODS. Permanent links to the content providers’ audio item online have been provided to enable online consultation. MP3s are being streamed online via a flash player.

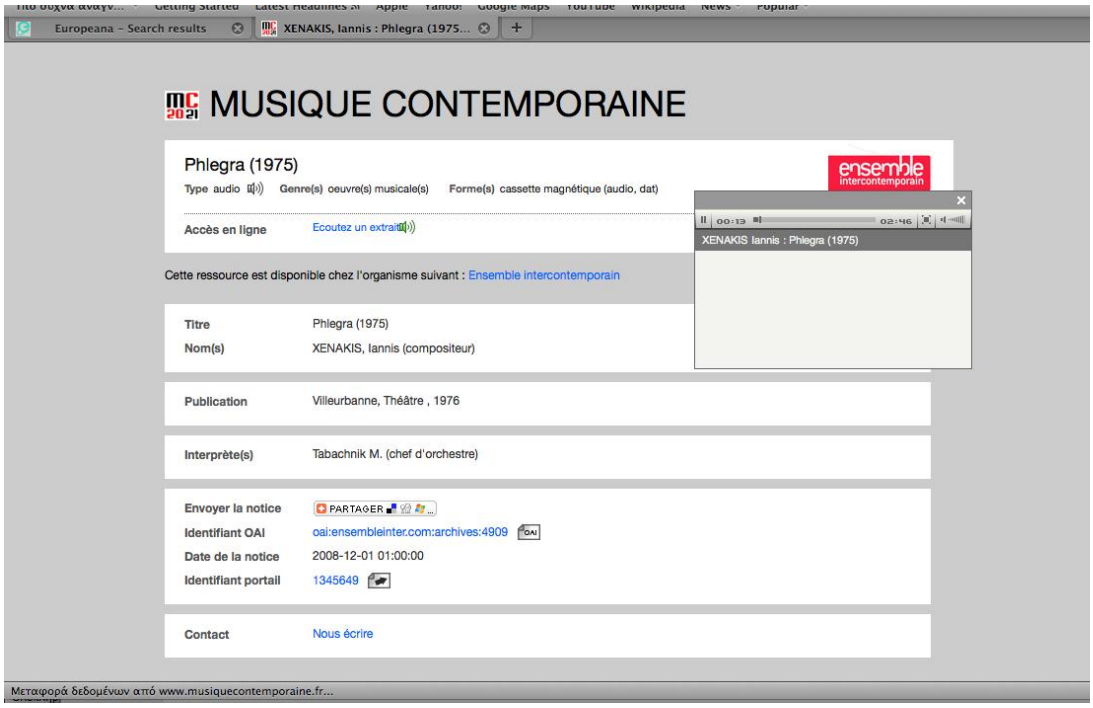

Figure 7: online consultation of a record in the Portal "Musique Contemporaine".

Another major AV content aggregator currently in Europeana is VideoActive with 8.553 videos. Videoactive (www.videoactive.eu) includes a significant collection of television programmes and stills from audiovisual archives across Europe. It also provides articles and comparative analysis on European TV history. VideoActive makes available resources to explore the development of television in Europe, but also the televising of cultural and historical events within and across nations. VideoActive's contents will be merged with EUscreen at the end of 2010 and we will present the latter in greater detail further on.

Other major AV content providers and aggregators in Europeana currently are: Het Geheugen van Nederland (890 videos), Mediatheque Valais-martigny (366 videos, 1.529 audio items), Cite de la musique (189 videos, 1.099 audio items), ABM-Ui TV ikling: 144 (videos), The British Library (1.565 audio items). Other providers have made available small collections of less than a hundred objects each (Biblioteca nacional de Portugal, Association des cinémathèques Européennes, IRCAM, etc).

4. Workflow for a new content provider

The current section describes the steps that a new content provider should follow in order to submit content to Europeana. The first part includes a questionnaire that helps a new content provider identify the best route to submit content to Europeana, either directly or through a known aggregator. The next part describes schematically the overall process of establishing a relationship and of submitting data to Europeana.

4.1 Routes to Content: helping a new organisation identify the best route to submit data

Europeana works directly with aggregators in order to harvest their data. Europeana frequently receives requests from local, regional or national institutions that wish to submit their data. These institutions are then guided, by the Europeana Office to the collaborating aggregators which will take care of the new content. However, if an institution cannot be redirected into an existing aggregator the Europeana Office will work directly with them. The structure and workflow related to directing new content providers is illustrated below:

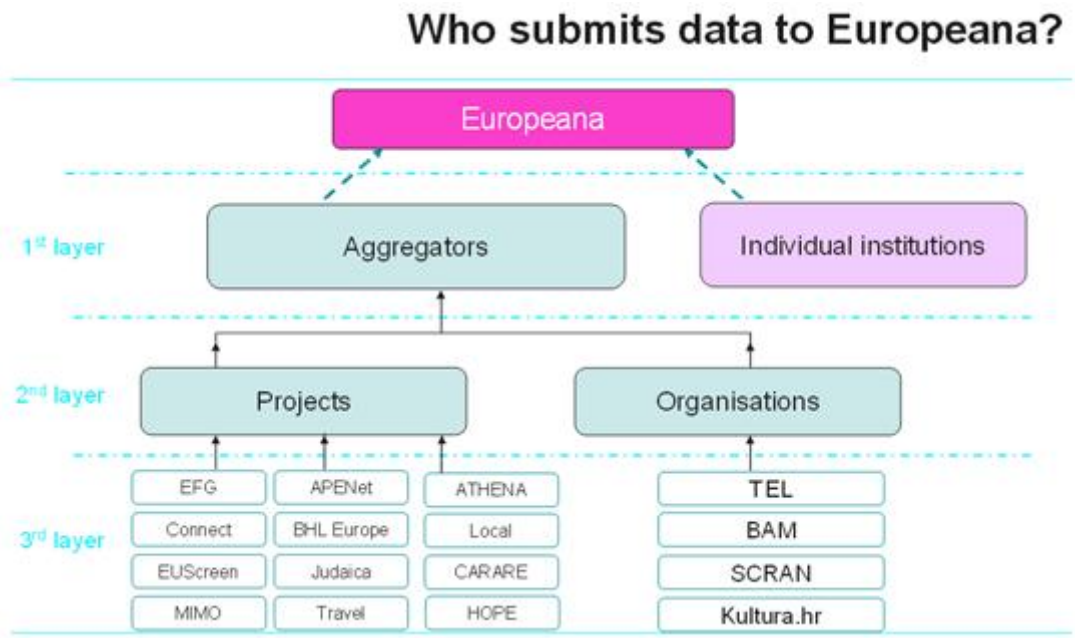

Figure 8: Submission of data to Europeana

New content providers are asked to submit the information described below using the online form in the section 'Routes to deliver content to Europeana': (http://version1.europeana.eu/web/guest/providing-content/). This helps Europeana to identify the appropriate route for a new content provider to submit content: this can be either via an existing partner aggregator or directly to Europeana.

The route to contribution of content to Europeana is determined by four main criteria:

a. Country
b. Sector
c. Scope
d. Geographic level

a. Country
National institutions/portals representing several sectors are the preferred first contact point for Europeana. A new content provider is, therefore, first directed to the national aggregator, if one is established. However, in some countries, these national institutions/portals are not yet established, and another routing will then be suggested.

b. Sector
In the current option the organisation selects the sector it is representing based on the sector-specific activities: Museum/Gallery, Library, Archive, Audio Visual, Research & Education, Portal, Publisher, Private, Others.

c. Scope
The criterion relates to the scope of the material for which you will be collecting data, whether as an aggregator defined by three domains, or as an individual organisation:

- Single aggregator: organisations representing and collecting data from one sector only, such as a museum, a library or an archive (for example, a national/regional/local museum, national/regional/local library, national/regional archive),
- Cross-domain aggregator: organisations working across several sectors and collecting material from museums, libraries and archives (for example, national portals and Europeana),

- Thematic aggregator: organisations bringing together content on a specific theme from one or several sectors (for example, Judaica that collects material from multiple sectors about the Jewish urban culture),
- Individual organisation**:** the organisation is submitting data only from its own holdings and is not acting as an aggregator.

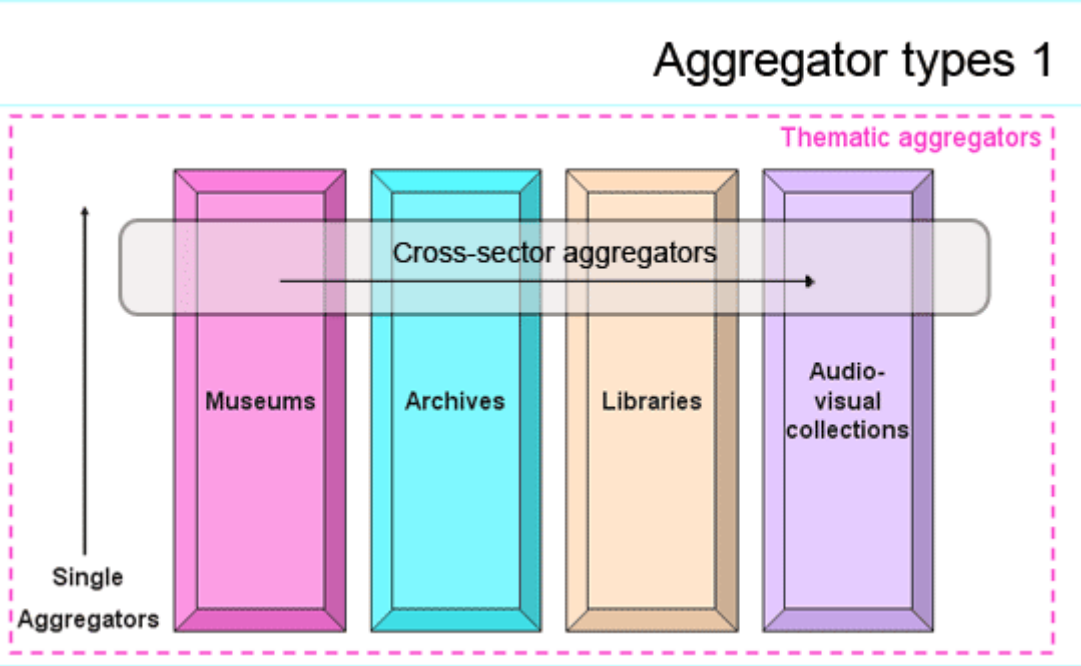

Figure 9: Aggregator types

d. Geographic
Geographic levels refer to the level at which an organisation operates in terms of the data it is collecting. Are the sources regional, European or worldwide? An example of an organisation operating at a European geographic level is The European Library.

Following the receipt of the above information, the Europeana Office informs by email the organisation whether the data will be ingested directly or via an existing partner aggregator. Individual aggregators have different procedures and requirements in place for accepting new content providers. If another aggregator is suggested, the organisation will make a direct contact with the aggregator and will be informed about its requirements for the submission of the data.

If an institution has problems completing the 'Routes to Content', a Partner Application Form can be completed and sent to the Europeana Office. Once the Partner Application Form (http://version1.europeana.eu/web/guest/providing-content) is processed by the Europeana Office, the office contacts the most suitable aggregator and consults with them on how to proceed further.

If the Europeana Office decides to aggregate the content directly, as in the case of aggregators, the organisation is asked to fill in the Partner Application Form. This is explained further in the following section.

Section 3.2 presents the Partner Application Form and section 3.3 describes the steps that an aggregator follows in order to submit data directly to Europeana.

4.2 Partner Application Form

The Partner Application Form allows Europeana to gather essential information about the aggregator and about the content it intends to submit. The Europeana Office processes this information in order to assess the workload and establish the next steps.

It is important that an aggregator provides a list of collections and organisations that it aggregates the content from. Currently, it is the only way for Europeana to identify and avoid potential duplication of identical content delivered via separate sources.

Once the aggregator submits the form, the Europeana Office will make contact within five working days to discuss the application. From then on, the aggregator will work with the Europeana Content Ingestion team to plan a timetable for the ingestion of its data.

The Partner Application Form should be returned as an email attachment to content@europeana.eu.

4.3 Data submission workflow

Any direct data submission to Europeana implies that the organisation follows the steps described here. Therefore this procedure also applies to organisations (including aggregators) that have already submitted data to Europeana in the past.

Figure 10: Direct data submission workflow

Step 1

The Aggregator and the Europeana Office formalise their collaboration by signing the Data Aggregator Agreement.

Step 2

The Europeana Content Ingestion team sends the Data Submission Form to the aggregator, through which it provides a description of the submission it wants to make: type of submission (new/update), licensing and metadata information related to this particular submission, transfer mechanism that will be used, etc.

Step 3

Within one month the Europeana Office reviews the submission request, gives feedback to the aggregator and arranges the planning for the submission of the data. Note: in some situations feedback from the Europeana Office can occur in a relatively short time after the reception of the request, especially in case of projects where a project-specific planning might impact the ordinary timing of the workflow's steps.

Step 4

In the next stage the aggregator executes the actions aiming at preparing the data sets to be submitted using the ESE specifications. The Mapping and Normalisation Guidelines are available to support the mapping exercise. The XML Schema is available to validate the mapping to ESE (note that these are explained in the next section).

In addition to the previous resources, the aggregator can use the Content Checker. The Content Checker is a web tool allowing the validation of mapping against the XML Schema, and to simulate the ingestion operations so that the search, browse and display of the data can be tested in a copy of the Europeana portal.

In parallel to the data preparation steps, the aggregator sets up and/or configures the data transfer infrastructure (in case of OAI-PMH, population of repository, aggregator in sets, etc.) and tests it.

Step 5

After the aggregator has tested the data sets using the Content Checker tool, the Europeana Office validates transfer methods with the aggregator.

Step 6

The Europeana Office runs the operations aimed at ingesting the data into the Europeana production environment. When the operations are completed, the aggregator receives a notification from the Europeana Office. Final steps after the data have been published on the Europeana portal include the aggregator's checking of their data in production to ensure that the data are presented in the desired way.

Throughout the process of submission the content provider or aggregator interacts with the following teams:

- The Europeana Business Development team is responsible for initial assessment of the Partner Application Form, assistance to the new content provider as well as the signing of contractual agreements (partnership, licensing).
- The Europeana Operations team is responsible for the ingestion of the data sets according to the information provided to them on the submission forms. They provide support and guidance and perform harvesting and quality checking of the data provided. They also give aggregators feedback about the data.
- Each project also has a liaison point at the Europeana Office to facilitate the communication and coordination with Europeana. The Europeana Projects Liaison team is responsible for the alignment of planning, identification of contacts and the transmission of key information on both sides.

4.4 Main aggregation routes of audiovisual content into Europeana

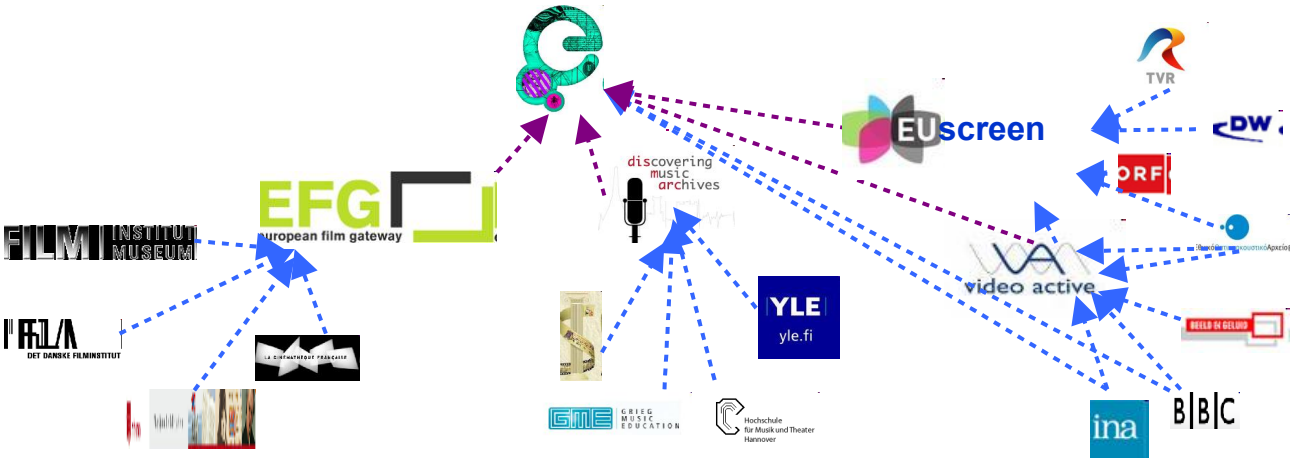

Figure 11 – Main aggregation routes of audiovisual content in Europeana

In the figure above we see the main routes for delivering AV content to Europeana. Some large organisations such as big broadcasters are likely to deliver content directly to Europeana. In the schema we also see the AV projects that aggregate content and will deliver it to Europeana as well as through their own portals. These aggregators have their own specifications with regards to how content is made available to them as we will explain in a following section. However, they deliver metadata to Europeana which is harmonised and normalised according to Europeana guidelines.

5. Technical and operational requirements for data submission

In the current section we present the technical and operational aspects of submitting data to Europeana and of making use of the Europeana open source code for developing tools and services.

5.1 Some basic principles

Every content provider and aggregator must comply with Europeana's technical requirements when submitting data. Before discussing these requirements, the basic principles upon which the technical guidelines are based on can be found below.

Europeana provides a common access point to digital cultural heritage objects across different cultural domains. It aims at complementing and not duplicating the initial environment of the digital objects. To achieve this, Europeana uses specialised portals. The object is linked to Europeana and is shown in a neutral environment, while at the same time still being available in the domain-specific context, which gives it a clearer definition.

Europeana aims at providing access to digital objects at the lowest possible level of granularity. This implies giving as direct as possible access to the digitised object itself, that is, with a minimum click distance between the description and the object. The minimum granularity can vary, and it is in the discretion of the content provider to decide this. A broadcast provider, for example, might decide to cut down a programme made of individual fragments, to make available each one as a separate digital object. On other occasions, the complete program is of value due to the context it adds to the individual fragments. Europeana asks that content providers keep the users in mind when deciding on the granularity of their data.

Europeana's data model enables search and discovery of digital objects. Europeana maintains a common central index of the objects' metadata. It has, therefore, an object-centric rather than a collection-centric approach.

Europeana stores *representations* of digital objects and not digital objects themselves. Europeana generates a description and a preview of digital objects with the help of the metadata and thumbnails or previews of the digital objects. This requires that on the content provider's/aggregator's side there is a repository, where the digital objects are stored and can be linked to. It also requires a native website that can be used to view, play and reuse the objects.

A digital object in Europeana is a unique single entity, which can be viewed/played by users (e.g., mpeg movie, mp3 audio, jpeg photo, PDF text, etc.) on their computers. A digital object is the digitised version of a physical/analogue cultural item/artefact. Europeana does not accept descriptions, even extensive ones, which do not lead to a digital object. A scanned catalogue card is, therefore, not a desired object for Europeana.

Europeana harvests, stores and indexes the metadata in a central index. This is important to enable integrated results display and processing of the data to provide various functionalities for the users (i.e. integrated and faceted search). This is also essential for the enrichment of the data. The preferred method for harvesting is the OAI-PMH protocol[19] and partner aggregators are required to set up an OAI-PMH repository comprising their data mapped to the Europeana Semantic Elements standard (see below).

Persistent Identifiers are of further importance for Europeana, because of the role they play in preventing duplication. Europeana has a wide network of aggregators and content providers, and

[19] The Open Archives Initiative- Protocol for Metadata Harvesting: http://www.openarchives.org/pmh/

the possibility of data being ingested more than once is present and should be avoided at any rate. The development and attribution of Persistent Identifiers will help de-duplicate content and provide greater control over the data.

5.2 What data are submitted?

A content provider or an aggregator is responsible for making available to Europeana the following data:

- Metadata (descriptive, administrative) describing a digital object. The metadata must be mapped to the Europeana Semantic Elements v3.2.2,
- A preview (thumbnail) of the described object,
- Active and stable links to the described digital object on the provider's site or the portals site.

The Europeana Semantic Elements specifications are explained in the next section.

With regards to thumbnails, Europeana strongly recommends the submission of a representative thumbnail image of each digital object in order to enhance the display of objects in the gallery of results. Ideally, an aggregator will be able to provide the URL of an existing thumbnail for each digital object. If there is no such thumbnail, Europeana can generate one using the link to the object provided in the <europeana:isShownBy> element. Not all image formats are suitable for this process, however, as the software used (ImageMagick) does not support everything. In addition, direct access to the image is needed to carry out this process; an image embedded in a web page is not suitable.

The Europeana data provider and aggregator agreements mentions audio and moving image previews. An audio preview is defined as a short, lower resolution extract of audio content. A moving image preview is a short, lower resolution extract of moving image content., which can be as short as one frame and functionally equivalent to a thumbnail. However, these specifications have not been turned into technical requirements. The current practice is that content providers provide a keyframe to be used as a preview for moving images. If this is not made available then a Europeana default type icon will be shown in the display. For audio items the providers either provide a representative picture, i.e. the logo of the collection or otherwise the Europeana default type icon can be displayed.

Europeana will be producing shortly a list of specifications for different formats for this purpose. In the meantime, more information about thumbnails and the metadata elements that support their use is given in the Europeana Mapping and Normalisation Guidelines.

Europeana creates only one thumbnail per record. If a record contains several pages, for example a PDF record, the front page is used to create the thumbnail. If a record contains several image files, the first of these is used for the thumbnail. If no suitable thumbnail can be obtained a default icon will appear in the portal.

The persistent link to a digital object and/or a full information page should be given as a URL linking directly to the digital object. If a link to a stand-alone object is not possible, a link to the object in a full information context can be provided. An example of this is an image embedded in the local web page. Care should be taken to ensure that the object is available for viewing with one click from the link provided.

Since the audio/video content is played on the provider's website, any format and any streaming solution is possible - it is up to the content provider to decide. It is up to the providers and aggregators to also determine video quality. Europeana only recommends that providers and aggregators should try to make the best quality possible available to the users.

5.3 Europeana Semantic Elements

There are two documents that data providers should consult prior to providing metadata to Europeana. These are the Europeana Semantic Elements Specification V3.2.2 (ESE) and the Metadata Mapping and Normalisation Guidelines V1.2. Both of these documents can be found on the Europeana Group web page (http://version1.europeana.eu/web/guest/technical-requirements/). Note that these documents are revised occasionally for clarification.

ESE is the metadata set developed for the prototype version of Europeana, launched in November 2008, and will be used for the Rhine release in July 2010. It is a Dublin Core-based application profile providing a generic set of DC elements and some locally-coined terms, which have been added specifically to support Europeana's functionalities. A full alphabetic declaration of these terms can be found in the ESE V3.2.2.

To provide metadata in the ESE format, contributors have to map their metadata to ESE. For assistance in this process, providers are recommended to look at the Metadata Mapping and Normalisation Guidelines; additional information can be also found in Section 5.5. In order to ensure machine readability, a normalisation process should be carried out on some values, such as the Europeana mandatory metadata fields. This information is described further in the Guidelines. There are also some general guidelines to consider when mapping. Providers are, for example, asked to consider how their data will perform in response to "who, what, where and when" queries. Content providers are strongly advised to consult the Guidelines before the submission of their data.

There are some mandatory elements in ESE. Most of the metadata is used to support the different search functions. However, it is important to supply as much information as possible, as this increases the visibility of the digital object in Europeana. If no usable data are provided, the object will not appear in the Timeline or in the Date facet.

Europeana currently handles four object types: text, image, sound and video. Content providers must map their local data to one of these Europeana type values. There is currently no category such as "museum object". An image of an object, for example, would be "IMAGE" and an image of an OCR document would be "TEXT." If the full content consists of moving images rather than a static image, the object type would be "VIDEO".

The following contents are not considered as videos in Europeana: a photograph taken during a video shot, a movie poster, a picture of a reel, etc…

With regards to submitting the links for the video/ audio content, the current are the following:

1. Player embedded in a web page. See example in figures 12a and 12b.

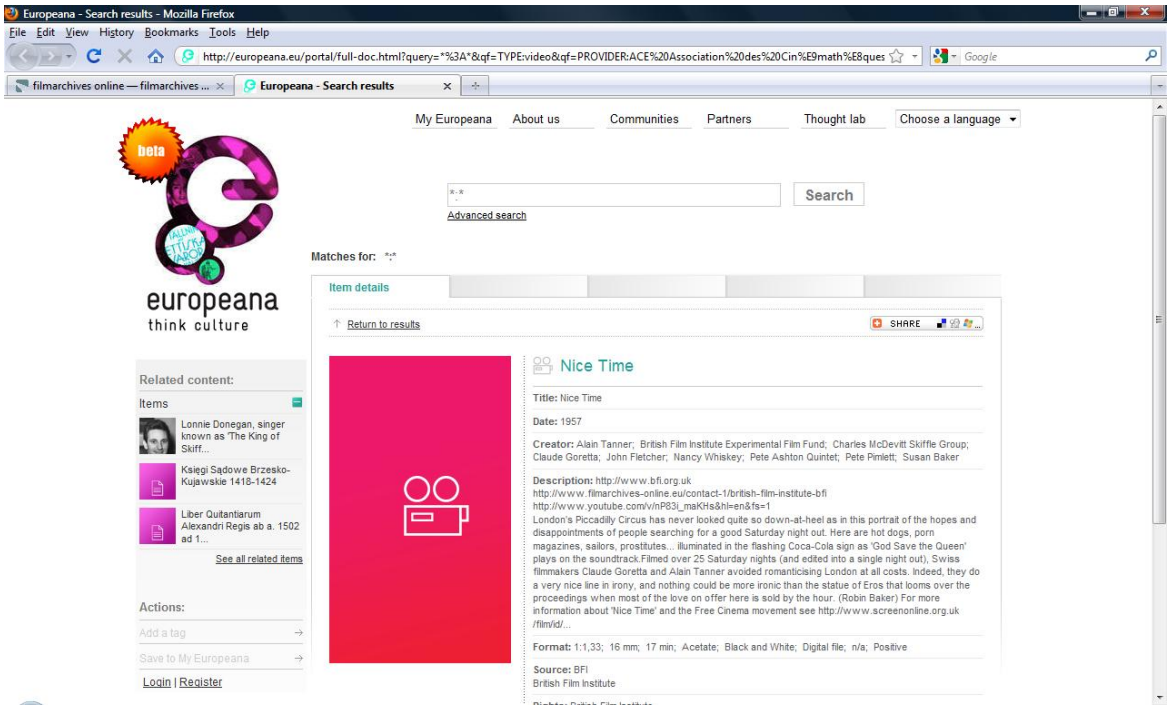

Figure 12a: Full result for an object with the Europeana Type "VIDEO". The user accesses the full video content by clicking on the thumbnail (see figure 12b).

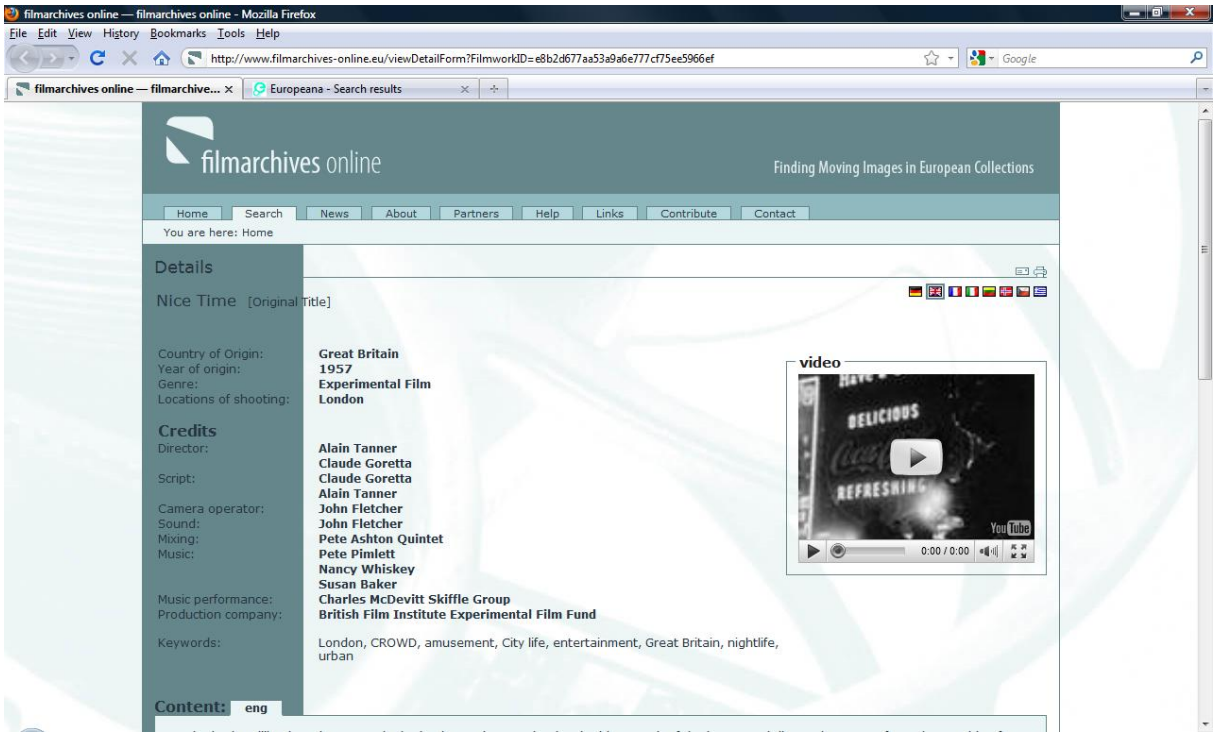

Figure12b: The Europeana user accesses the full video content on the provider's website. In this case the video is played in a player embedded in the page

Note: The URL to such page would most probably be held in the Europeana metadata element *isShownAt* since it is likely to include descriptive and navigational information next to the video content.

2. Player resolved through a URL supplied by the Provider or Aggregator.

Note: The URL to a player should be held in the Europeana metadata element *isShownBy* since it is likely NOT to include descriptive and navigational information next to the video content.

Note: The URLs provided via the *isShownBy/isShownAt* metadata elements to link to the video/audio content are currently of the following types:

1. http://...
2. mms://...

Other syntax might be accepted but it will need to be tested first by Europeana.

For the current version of the ESE Specifications v.3.2.2 visit: http://version1.europeana.eu/c/document_library/get_file?uuid=c56f82a4-8191-42fa-9379-4d5ff8c4ff75&groupId=10602

5.4 Metadata Harmonisation and Normalisation Guidelines

In addition to the mapping of the metadata, it is necessary to carry out a normalisation process on some values to enable machine-readability. This is particularly true of some of the Europeana terms, which are designated as mandatory and must contain values in standard form. This information is given in the Guidelines, which go into more detail about mapping source data to the ESE format.

For the Metadata Mapping and Normalisation Guidelines visit: http://version1.europeana.eu/c/document_library/get_file?uuid=58e2b828-b5f3-4fe0-aa46-3dcbc0a2a1f0&groupId=10602

5.5 Content validation

Content providers and aggregators submitting data directly to Europeana have two tools at their disposal to test and validate the datasets they have prepared for compliancy to ESE: the ESE v3.2 XML Schema and the Content Checker.

The ESE v3.2 XML Schema[20] is the XML representation of the Europeana Semantic Elements (ESE) specifications v3.2.This schema can be used to validate XML instances of Data Sets to be submitted to Europeana.

The Europeana Content Checker consists of the Content Checker Ingestor and the Content Checker Portal. The Content Checker Ingestor is for data providers and aggregators to ingest data into the Europeana test and validation environment. The application allows content providers to upload, test compliancy with ESE, index records as well as cache thumbnails. The Content Checker portal enables data providers and aggregators to search and browse their data in a simulated Europeana portal.

Once this process is complete and content providers and the Europeana Office agree on the state of the datasets, the Europeana Office starts the actual process of harvesting and indexing.

[20]You can read about the XML schema http://version1.europeana.eu/c/document_library/get_file?uuid=104614b7-1ef3-4313-9578-59da844e732f&groupId=10602 and you can find it here: http://www.europeana.eu/schemas/ese/ESE-V3.2.xsd

6. Legal Requirements

6.1 Data Aggregator and Data Provider Agreements

In order to make their data available content aggregators as well as individual organisations that submit data directly to Europeana need to sign respectively a Data Aggregators or Data Providers Agreement with Europeana. The agreements grant the right to Europeana for use of metadata made available in order to produce the Europeana.eu data. The agreements define the roles and responsibilities between the signing parties.

Following a long consultation process with stakeholders, the final document has been drafted and includes a non-commercial use regime, both for Europeana and the third parties.

The term of the Agreement is set to one year and is automatically renewed for a period of one year on January 1 every year, unless terminated by one of the parties. The agreements define obligations of Europeana and the data aggregators.

The Agreement does not cover the content; it relates just to metadata. Europeana drives traffic directly to the data aggregator's site, where users can view the content in its original context, under the terms the aggregator chooses.

The priority is to keep implementation and transaction costs to a minimum for the data providers and especially the data aggregators. A streamlined mechanism for handling rights is necessary, and the terms of the Agreements have to be simple to implement in order to scale up to, potentially, many hundreds of providers.

The Data Aggregator Agreement is based on a clean-hands model. That means Europeana does not sign agreements with individual content providers submitting data via an aggregator. The aggregator is responsible, if necessary, to clear rights with individual content providers in order to meet the requirements of the Data Aggregator Agreement with Europeana. The aggregator can use the Europeana Content Provider Agreement as a model in this situation.

As well as the Agreement which governs Europeana's use of the metadata, the Europeana Office is also drafting a license to govern third party use.

The Europeana Foundation cannot grant more rights to third parties than those conferred by the Agreements. The license currently being drafted for third parties will be similar in spirit to a Creative Commons license but will be more robust, because the Creative Commons is not enforceable if data is not copyrighted. The license will be a Europeana-Attribution - Share Alike – Non-commercial license, and further information will be circulated shortly.

7. European audiovisual projects and their specifications

In the following section we will briefly present the main projects that Europeana collaborates with and which focus at the aggregation of audio and video content. The European Film Gateway is the common access point of the national film archives and Cinémathèques in Europe, EUscreen is the portal for broadcast and TV archives and the EuropeanaConnect Audio Network (aka DISMARC) is an aggregator for music and sound archives in Europe.

7.1 The European Film Gateway

7.1.1 Introduction

Initiated by the Association des Cinémathèques Européennes (ACE) and the Europeana Foundation, the European Film Gateway (EFG) project develops an online portal, which will provide unprecedented access to more than 700.000 digital objects. This unique collection includes feature and short films, images, posters, film set drawings, photos, sound material and text documents. The EFG consortium unites 21 partners in a ground-breaking alliance to deliver technology and content, among them 15 European film archives and cinémathèques.

EFG addresses a number of key issues for access to digital content, namely, technical and semantic interoperability, metadata standards as well as best practices for rights' clearance and IPR management of cinematographic works.

The Deutsches Filminstitut – DIF e.V. co-ordinates the project co-funded by the Community programme eContent*plus*. The EFG project started in September 2008 and will end in 2011.

7.1.2 Expected Results by the end of the project

- A single access point to the digitised collections of Europe's film archives
- Agreement on common interoperability standards for film archives and cinémathèques
- Best practices for rights' clearance and IPR management of cinematographic works
- Access to Europe's film heritage through Europeana.

Key deliverables to date:

- Common Interoperability schema for archival resources and filmographic descriptions, report on common interoperability schema – D2.3 – April 2009 http://www.europeanfilmgateway.eu/downloads/D22_Common_Interoperability_Schema_final_090904.pdf

- Basic guidelines and recommendations on copyright clearance and IPR management - D 5.2 – August 2009

7.1.3 Network strategy

Most of the film archives and Cinémathèques represented in EFG are part of the Association des Cinémathèques Européennes (ACE), an organisation aiming at protecting the European film

heritage and to assure that the audio-visual records of our century survive for future generations. ACE is an affiliation of 39 national and regional preservation film archives from all over Europe.

Members of ACE are also part of the “Fédération Internationale des Archives du Film” (FIAF), an organisation founded in 1939 for promoting the preservation of film as art and historical document, and for facilitating cooperation among its members. Membership is limited to non-profit film archives (governmental or non-governmental) which operate on a national level. Currently there are some ninety members of FIAF from Europe, North America, Asia, Africa, and Latin America.

FIAF’s main interests centre on the professional work of film archivists. Areas requiring highly specialised knowledge and expertise include cataloguing, documentation, and preservation.

The existing EFG partners are the following:

- Association des Cinémathèques Européennes, Frankfurt/Brussels
- CNR-ISTI, Pisa
- EDL Foundation, The Hague
- Eremo srl, Cupramontana
- reelport GmbH, Cologne
- FernUniversität Hagen

Film Archives/Cinémathèques

- Cinecittà Luce S.p.A., Rome
- Cinemateca Portuguesa, Museu do cinema, Lisbon
- Cineteca del Comune di Bologna
- Det Danske Filminstitut, Copenhagen
- Deutsches Filminstitut - DIF e.V., Frankfurt
- Filmarchiv Austria, Vienna
- Kansallinen Audiovisuaalinen Arkisto, Helsinki
- La Cinémathèque Française, Paris
- Lichtspiel – Kinemathek Bern
- Lietuvos Centrinis Valstybés Archyvas, Vilnius
- Magyar Nemzeti Filmarchívum, Budapest
- Národní Filmový Archiv, Prague
- Nasjonalbiblioteket, Oslo
- Nederlands Filmmuseum, Amsterdam
- Tainiothiki tis Ellados, Athens

7.1.4 Type of content

EFG, however, is not a digitisation project; the project focuses on access to already existing digitised collections. It is also not a video-on-demand project, since it has a scope that goes beyond the moving images: EFG will provide access to film-related material, if possible in its free of charge form. This implies a single access point to moving images and cinema-related material in Europe's film archives and cinémathèques. In other words, EFG will give access to:

• 65 individual collections,

• Several thousand hours of films and film clips including documentaries, film features, newsreels, shorts, experimental films, etc.

• Images including film stills, posters, photos, set drawings etc.,

• Text archive documents containing scripts, programmes, periodicals, censorship documents.

7.1.5 EFG Data format

In order to describe objects, the EFG metadata model basically distinguishes three levels: Creation, Manifestation and Item. In doing so, it takes into account considerations of the IFLA Functional Specifications for Bibliographic Records (FRBR), the Encoded Archival Context (EAC) Working Group and the initiative Metadata Standardisation for Cinematographic Works (CEN.BT TC372).

- "Creation" is the topmost level of description and refers to the concept of Cinematographic Work. A creation in EFG can be either an audiovisual work (represented in the "AVCreation" entity) or a non-audiovisual work (represented in the "NonAVCreation" entity - Non-audiovisual creations in EFG can be pictures, photos, correspondence, books or periodicals).

- This also applies to the "Manifestation" level which is represented in the entities "AVManifestation" and "NonAVManifestation" (Manifestation groups all properties that can change during the lifecycle of an audiovisual creation without affecting the identity of a film work as such. A manifestation is the physical embodiment of an audiovisual creation. Examples are archival copies (analogue or digital) or distribution files).

- The "Item" entity functions as a logical wrapper for the digital file presented in the EFG portal. The item entity points to the digital file held in the source repository. Since the item is not a part of the EFG database the Item entity will always contain a reference to an outside resource, either for immediate viewing or for ordering from an archive or distributor. The item functions as a logical wrapper for the digital object displayed in EFG.

- 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 elements: Name (which includes one or more Parts, such as the prefix, the forename, family name, etc., it also includes the geographic and temporal scope), Date (which specifies the temporal properties of the person in relation with his activity), Place (where the activity was performed), Sex, Type of Activity. Similar elements are defined for Corporate Body and Group.

- The "Event" is defined as a primary 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).

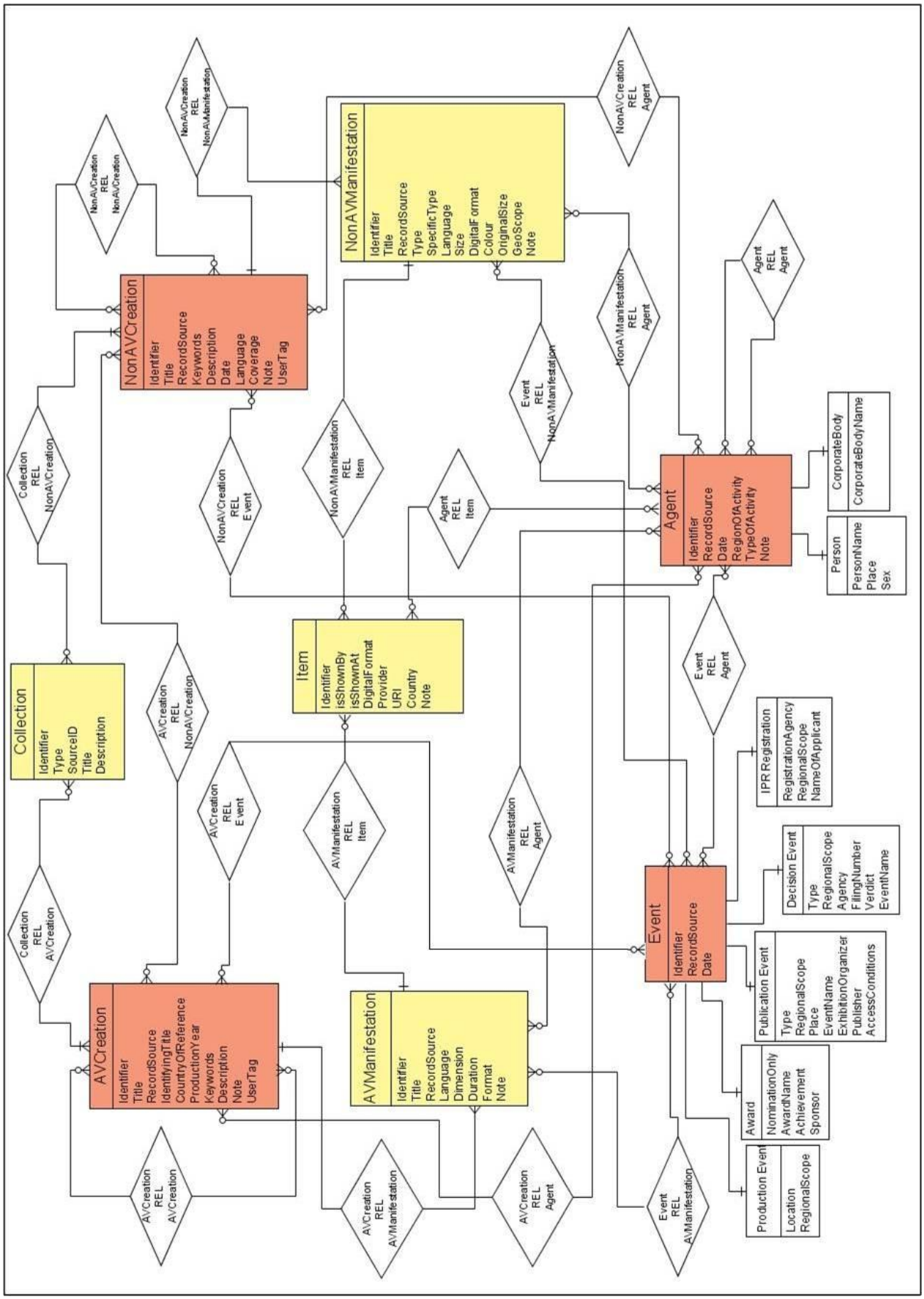

Figure 13: Entities and relationships of the EFG schema

7.1.6 Content ingestion workflow and tools

In general, making available metadata using the EFG XML schema as the exchange format will be preferred. However, alternative delivery formats are also possible. Several means of data transfer such as HTTP, FTP, OAI, SOAP are generally possible. In practice, HTTP, FTP and OAI have already been used for delivering metadata during the current beta phase of the EFG back-end system.

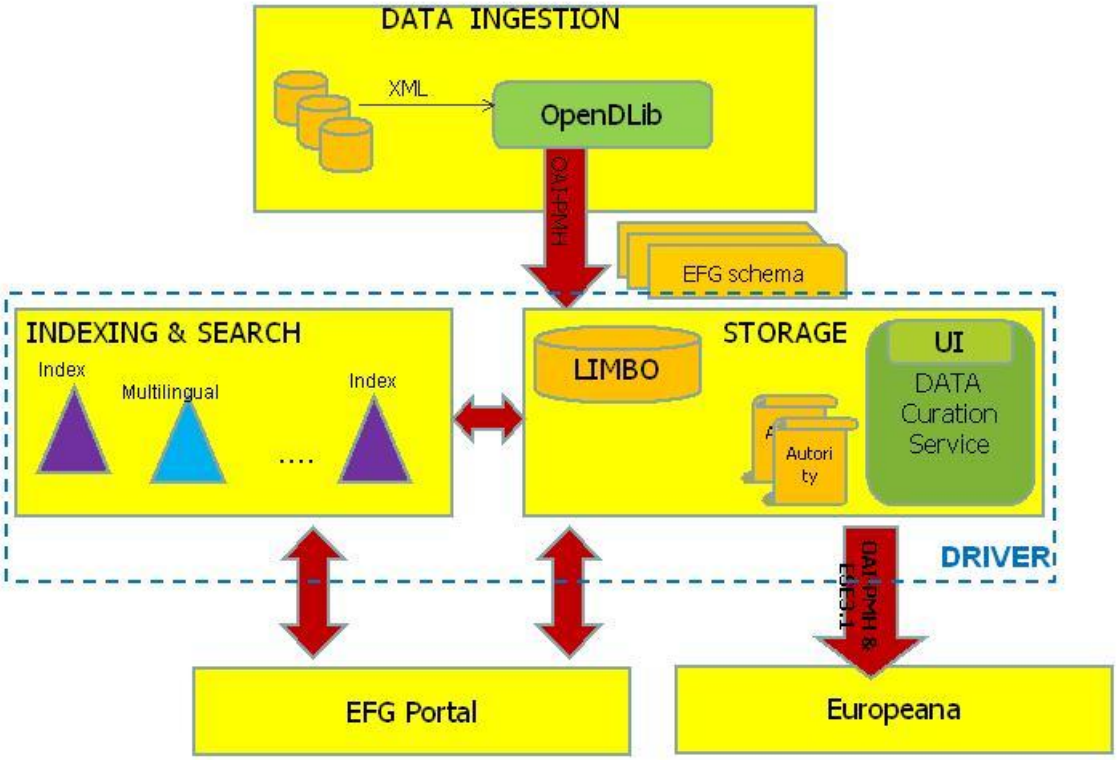

Figure 14: **EFG back-end architecture**

The DRIVER[21] system used to ingest, store, manage, index and search metadata contributed by the archives depends on OAI-PMH[22]. To enable also other forms of metadata provision apart from OAI-PMH, an intermediate ingestion step was introduced, making use of the OpenDLib[23] system. OpenDLib gathers metadata in various different XML formats, stores the exports and makes them OAI-PMH harvestable. This pre-processing may be skipped if a data provider already offers OAI-PMH.

EFG's guidelines for the inclusion of new data providers are scheduled to be published in February 2011 after a stable version of the EFG infrastructure has been achieved and after extensive experiences with metadata ingestion, transformation and management have been made. However, a basic and "hands-on" practical guidelines document was already issued to EFG partner archives: Preparing Metadata for XML Exports. Guidelines for EFG Content Providers, available on the

[21] http://www.driver-repository.eu/

[22] http://www.theeuropeanlibrary.org/portal/organisation/cooperation/telplus/documents/TELplus_D2.1_31052008.pdf

[23] http://www.opendlib.com/home.html

Guidelines and Standards section[24] on the project website. It specifies several basic requirements for XML metadata exports to be provided to EFG.

A data curation service is provided, which supplies the contributing archives with the possibility of editing metadata already ingested into EFG. In general, editing of metadata should be made in the archives' local databases. But there are several important exceptions to this rule, particularly concerning matching and merging film titles and person names. A User Interface enabling the archive's data curator to identify, match and merge records of different provenience describing identical entities (mainly persons and works of film) is also provided by EFG within the DRIVER system's infrastructure. It is supposed to minimize double entries and enable a reliable search for digital objects related to specific films or persons.

Data contribution to Europeana is done again via OAI-PMH, using the Europeana Semantic Elements V3.2. This means that a considerable amount of contextualized information which is gathered and kept in the EFG-system is not used within Europeana. This includes for example further information about works of film or persons, and especially relationships between the two.

7.1.7 EFG portal specifications

The European Film Gateway project also implies the conceptualisation and implementation of a Web portal providing search functionalities and access to digital objects from European film archives and cinémathèques. The EFG portal is planned to be launched in Autumn 2010.

The EFG portal will provide information about the digital objects, derived from the metadata as supplied by the individual film archive, along with additional information about films and people associated with these films (filmographic information) also derived from the individual archives.

Digital objects from the film archives will be mainly photos, but also videos, text documents and sound files. They will be stored locally on several servers, usually within the archive institutions and will be made available through institutions' websites and digital repositories. Consequently, the content itself will not be physically copied to the central EFG server. Only the metadata will be copied to the central EFG server and ingested in the EFG Information Space. Among other information, the metadata will contain the direct URL to the local file for embedding the image/video on the EFG site through deep linking.

Several ways of discovering and presenting the digital objects should be implemented, such as simple search, advanced search, filtering search results using various categories and "exploring" the collections through browsing. Additionally, the portal should allow user interaction, such as adding user tags / keywords for displayed objects. It should allow bookmaking maintenance in a personalised workspace ("Community").

As EFG is an international project, the front-end of the website must be displayed in all languages supported by the project. At the moment these are the following: English, French, Italian, Portuguese, Danish, German, Lithuanian, Hungarian, Norwegian, Czech, Dutch, Greek.

7.1.8 EFG and Europeana

The European Film Gateway will be linked to Europeana, the cultural website governed by the Europeana Foundation. By contributing to Europeana, EFG will add to making Europe's diverse cultural heritage available to a wider public via a single access point. EFG unlocks the richness of Europe's film archives for the Europeana users and cinema lovers all over the world.

[24] http://www.europeanfilmgateway.eu/guidelines_and_standards.php

On a technical level, metadata from EFG is contributed to Europeana via OAI-PMH using the Europeana Semantic Elements. This entails a “dumbing-down” of the information to some degree. Taking this into account, an own EFG Portal is also being designed and developed within the framework of the project. This portal provides more detailed information about the objects and especially about the relations between digital objects and “filmographic information”, i.e.: data about works of film and persons working or having worked in the field of cinema and film. The EFG Portal in turn also provides the links to the digital objects in their original context, such as www.filmportal.de.

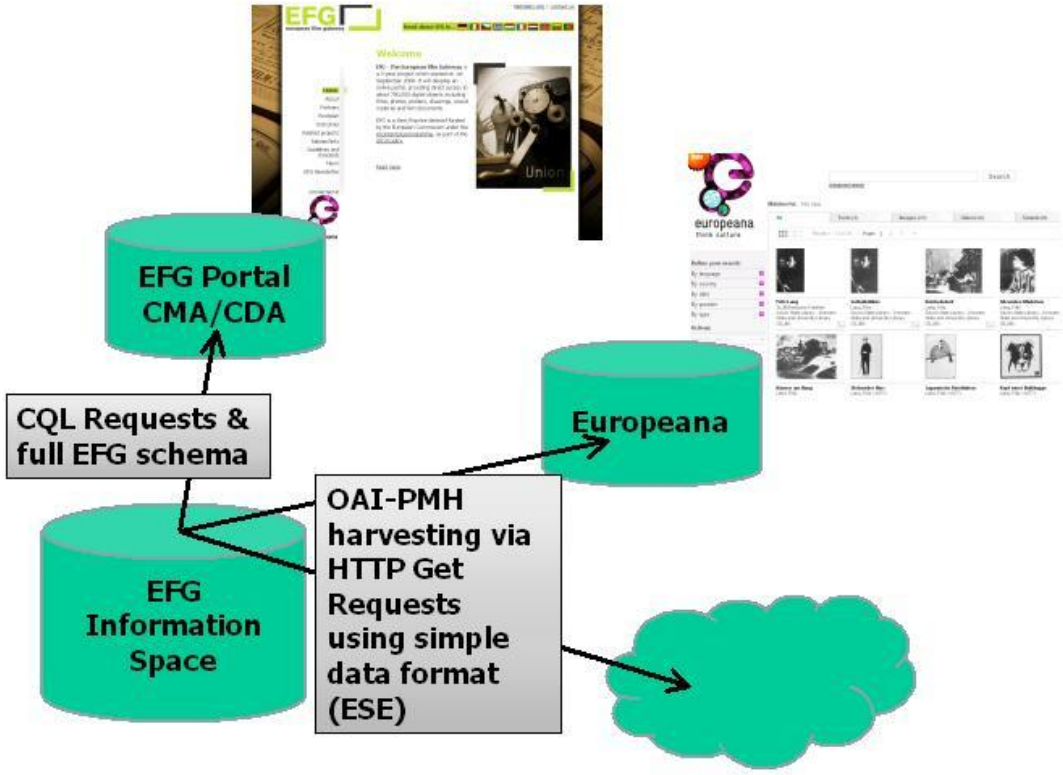

Figure 15: EFG portal architecture

7.2 EUscreen

7.2.1 Introduction

EUscreen is another initiative in the European AV domain aiming at enhancing access to mainly TV content coming from broadcasters’ archives. It was initiated as a Best Practice Network funded by the eContentPlus programme and is supported by FIAT/IFTA, the European Broadcasting Union and the EDL Foundation.

The particular issues the project addresses are technical and semantic interoperability across TV archives, complex rights issues involved with opening access to AV content online and contextualization of the content to make it more relevant for users’ particular needs.

EUscreen is a three-year project that began in October 2009 and is coordinated by the University of Utrecht. The project will make available via a dedicated multilingual portal and Europeana a core

collection of more than 35,000 items like videos, articles and photographs online by the end of the project. The first version of the EUscreen portal will be released at the end of 2010.

The EUscreen project builds on the experience of a consortium of partners who worked on the EC funded Video Active project[25]. Running from 2006 – 2009 Video Active created multi-lingual access to over 9,000 items of audio-visual content (predominantly television content) from broadcasters and archives around Europe. The project, funded out of the *e*Content*plus* programme, included eighteen partners from twelve countries and involved eleven languages..

Since the EUscreen project only started in October 2009 much of the information here describes planned and proposed activities.

The following results are expected from the project over its' 3 year lifetime:

- A highly interoperable EUscreen platform allowing the multicultural and multilingual exploration of content and metadata from 19 EU member states, integrated into and harmonised with Europeana so that Europeana can access the EUscreen semantic resource seamlessly.
- The EUscreen core collection of European television heritage. A publicly available collection of >35.000 items from television programming of the consortium content providers throughout Europe.
- Additional information linked to the core collection of European television heritage.
- Three conferences and nine workshops plus proceeding reports focusing on specific creative and strategic topics that stakeholders are facing (such as metadata standards, IPR solutions, digitisation policies and new services and business models).
- Nine well defined and tested use cases in the fields of research, learning and leisure (including academics, school pupils, b2b, cultural industries and the cultural heritage sector) as well for the benefit of open culture productions.
- An e-journal dedicated to the study of European television based on exploration of the archives and of EUscreen content in particular.
- A lively community that actively seeks to engage new members and liaise with existing networks.

7.2.2 Network and partner strategy

EUscreen is co-ordinated by The University of Utrecht. The EUscreen consortium consists of 27 core partners and 8 associate partners from 17 EU member states plus Switzerland. Besides the widespread coverage of content providers there is an enormous potential for networking and creating impact through the collaboration with EBU, FIAT/IFTA, the European Digital Library Foundation and other AV, educational and cultural heritage networks.

The current partners to the project are:

Archives- content providers

- British Universities Film & Video Council
- Deutsche Welle
- Danske Radio
- Hellenic National Audiovisal Archive
- Institut National de l'Audiovisuel
- Cinecittà Luce

[25] http://videoactive.eu/VideoActive/Home.do

- Osterreichische Rundfunk
- Radiotelevisione Italiana
- Radio-Télévision Belge de la Communauté Française
- Raidió Teilifís Éirann
- Radiotelevizija Slovenija
- Kungliga Biblioteket
- Televisió de Catalunya
- Telewizja Polska
- Televiziunea Româna
- Vlaamse Radio & Televisie
- Ceská Televize
- Nederlands Insituut voor Beeld en Geluid

Technology providers

- European Broadcasting Union
- Noterik

Europeana

- Europeana Foundation

Research organisations

- ATiT
- ELTE
- National Technical University of Athens
- Royal Holloway University of London
- TAIK
- Universiteit Utrecht

Associate partners

- BBC
- FIAT/IFTA
- AthenaWeb
- AAMOD
- Politechnico di Torino
- Audiovisual library of the EC
- DIVERSE
- Memoriav

EUscreen aims at building a long-term sustainable service to maintain and expand access and to build further on the project's goals after the end of the project. The underlying business model to include more partners and the sustainability aspects will be investigated over the project's lifetime. Before the end of the project EUscreen foundation will be setup with the aim to sustain the goals set in the project in the future.

7.2.3 Types of content and content strategy

Over the project's lifetime, EUscreen partners will make available a core collection of >35,000 television items as well as references to digitised items of the institutional collections and catalogue entries.

The Royal Holloway University of London leads the task of defining the content selection policy, the common metadata scheme and strategies for content contextualisation. The content selection and metadata scheme is being developed in consultation with the EUscreen content providers and the official content selection policy and metadata scheme will be announced in June 2010.

Content selection will be based on a review and most likely adjustment of the Video Active content selection policy as well as on user-led demands and interests. The selection of content in Video Active was motivated by a core principle which was to compare programming that reflected the cultural and historical similarities and differences of television from across the European Union. The content also provided an important record of key social, cultural and political events and developments in European history. So Video Active had at its heart the twin concerns of the history of television in Europe and the history of Europe on television.

For EUscreen this is still a valid approach as the selection of content, metadata and contextual analysis will facilitate and reward comparative investigation and study. Yet the EUscreen content selection policy will need to take account in its definition the following:

- developments in Intellectual Property Rights, preservation strategies and digitisation policies that relate to access to metadata and content in each partner country as well as intended integration with Europeana. Given the national digitisation policies and the prevailing IPR practices and framework in each country, it is likely that a content selection policy will firstly be oriented towards factual programmes to provide (cultural) information to the broadest range of EUscreen and Europeana users; however the content selection policy will also consider models for inclusion of non-factual and entertainment forms of content where holdings and IPR allow (using the experience with Video Active).
- the various scenarios of use of audiovisual content in research, learning and leisure. EUscreen places particular importance in the development of scenarios for using this content in different contexts. The definition, therefore, of target groups and their particular needs is likely to have a significant impact in the project's content selection as well as data model specifications.

The EUscreen content selection policy to be launched in June 2010 will detail the themes, topics and content to be selected in the remaining period of the project's current funding (until August 2012).

In order to make archive content understandable and meaningful, all content will be accompanied by metadata which will provide a baseline understanding of the provenance and description of its original source and how it appeared (for example, transmission date and original television transmission channel – where available).

Content providers will also be encouraged to provide other contextualising information such as script material; lost material such as live and unrecorded introductions; information and documentation from other relevant broadcasting stakeholders such regulators or trade unions. The complex and time-consuming workflows involved in sourcing and generating such contextual information means that it is likely that content providers will only be able to support a small amount of content in this way (for example, in areas where they are allowed curate their own virtual exhibitions.

Users will also be invited to contribute further material (e.g. reminiscences of watching television programmes, or working on its production, links to press articles, books, and existing course materials). Solutions for contextualisation will thus be proposed through the development of editorial mechanisms and the development of user-led activities such as rating and tagging systems, blogs etc, to support use of programme content. This is currently under discussion.

7.2.4 EUscreen data format

In order to achieve metadata interoperability between the TV archives, an agreed metadata format will be defined in the project. For the definition of the EUscreen metadata schema, the Video Active and the EBU Core Set for Metadata (EBU's instantiation of DCcore adapted to suit better television material) are being investigated and extended where needed to cover the representational needs of EUscreen use cases.

The EBU Core Set for Metadata has been identified as being the minimum information needed to describe radio and television content. The EBU core set of metadata is an extension of the Dublin Core. It is a flat list of attributes characterising a resource. An XML representation is provided in case this metadata set would be implemented in archive exchange projects using the Open Archive Initiative's Protocol for Metadata Harvesting (OAI-PMH). The Dublin Core is being used as core metadata by librarians and in cultural heritage projects with which radio and television archives have a natural link. The core set of metadata, used for such archives, offers a bridge between cultural heritage databases, broadcasting production systems and broadcasting archive repositories. Having EBU core as the core set of metadata, a solution for interoperability with Europeana's current ESE model is also provided. The EUscreen data format is currently under development and is expected to be finalized in June 2010.

Finally, When all the metadata are collected, they will be transformed in a Semantic Web language (i.e. OWL/RDF) and stored in a semantic metadata store (i.e. RDF store), they will be harmonized and enriched with the project's language resources and specifications. They will be aligned with and exported in the IEEE Learning Object Metadata (LOM) format as well in any existing (semantic) web format (i.e., XML, OWL/RDF, SKOS).

7.2.5 Content ingestion architecture, workflow and tools

The EUscreen system comprises of the Back-end and the Frond-end tools. The Backend tool is responsible for the annotation of the digital items and the Front-end for retrieving and presenting the metadata along with the video play-out service. In the current section we present the envisaged back-end architecture, workflow and tools. The architecture is schematically presented in the diagram below.

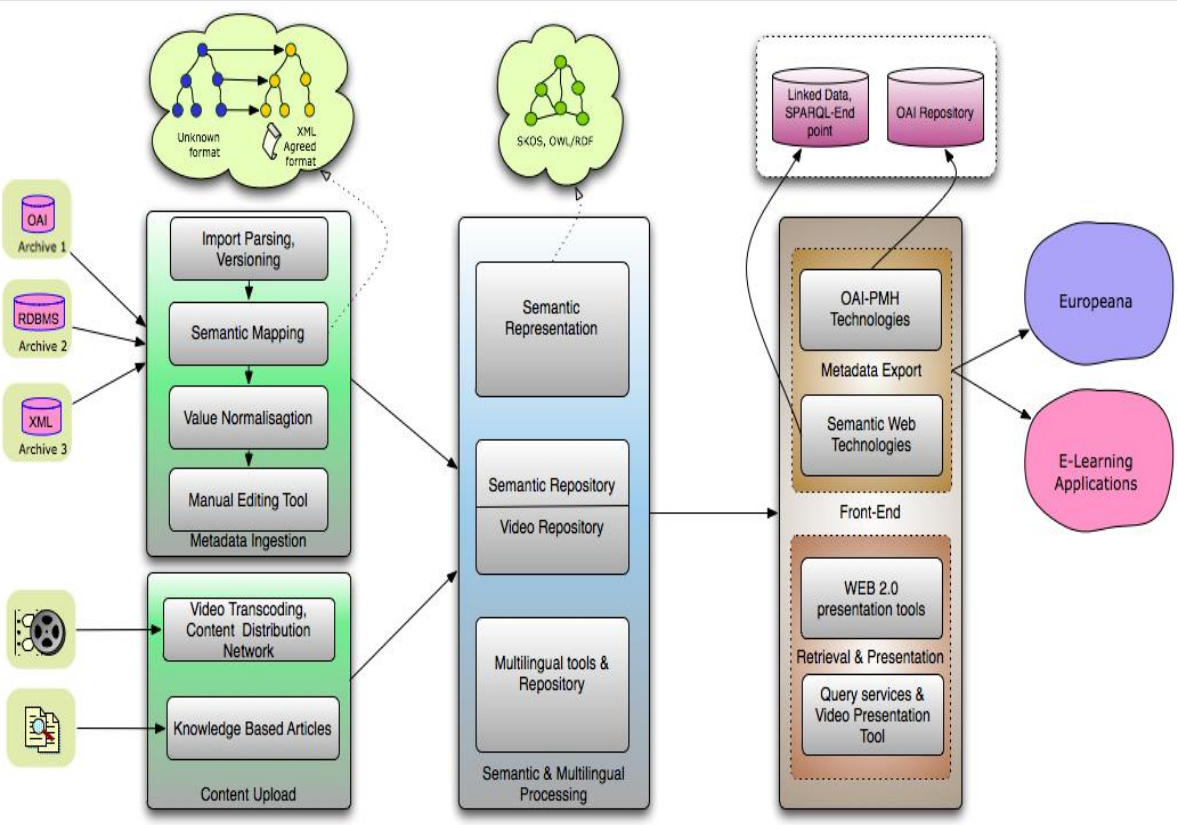

figure 16: Schematic representation of the EUscreen back-end architecture

EUscreen will be able to provide the link to the video items of content available on the websites of individual archives through OAI-PMH or alternatively by entering links in the EUscreen Metadata Enrichment Tool.

The archives providing content and metadata in the EUscreen application are represented at the left of the image above. Each archive will sustain the local (proprietary) metadata schema and will either set up an OAI-PMH repository to make available the metadata in the central EUscreen application or use the manual way of annotating and inserting the metadata in the semantic store through the use of the Manual Annotation web service. This annotation service will also include the possibility to import xml and Excel formats, extracted from the legacy databases of the partners. Partners will be responsible for mapping between their legacy system and the EUscreen data format.

When all the metadata are collected, they will be transformed in a Semantic Web language (i.e. OWL/RDF) and stored in a semantic metadata store (i.e. RDF store), they will be harmonized and enriched with the project's language resources and specifications.

The metadata will be exported in other metadata schemas and in various language formats aiming at making available the EUscreen metadata in e-learning, leisure and research applications. In particular, the EUscreen metadata will be aligned with and exported in the IEEE Learning Object Metadata (LOM) format. Additionally, the metadata will be exported in any existing (semantic) web format (i.e., XML, OWL/RDF, SKOS). Additionally, EUscreen metadata will be mapped to Europeana data specifications.

7.2.6 Multilinguality and language resources used

Multilingual search enables the user to enter a search query in one of 13 languages, and receive results with metadata from another language. The EUscreen platform will make use of controlled vocabularies to support multilingual search. A third party software component, Thesaurix[26], will be used to store the controlled vocabulary. Each of the content providers will be responsible for the translations of all terms into the different languages.

In the VideoActive project, the same software was used, and currently it contains around 1500 terms in eleven different languages.

7.2.7 Online accessibility of content – EUscreen front-end specifications

Metadata serve as a centralised starting point to access the archive. Once a user looks at the detail of an asset, he/she not only gets flat text metadata, but also a glimpse of the actual content of the video by means of a key frame extraction, provided during the ingest of metadata by the archive. The integrated Multilingual Thesaurus enables users to browse through collections based on available themes. Users are not only able to enter a flat text search, but are able to graphically search and browse the archive through the EUscreen portal.

In order to make archive content understandable and meaningful, all content will be accompanied by detailed descriptions of its original source and how it appeared (for example, TV channel, programme and schedule details; script material; lost material such as live and unrecorded introductions; information from other related broadcasting institutions such as regulators and trade unions). Users will be invited to contribute further material (e.g. reminiscences of seeing programmes at the time, or working on particular productions).

In addition to the aforementioned functionalities, the EUscreen portal will provide novel browsing and viewing functionalities. The EUscreen portal will feature a timeline-based search, which will provide the user with a temporal search experience. All assets that contain temporal information, such as transmission date, will be inserted into a timeline. Users can enter search queries and use filters to limit the amount of results that are shown on the timeline.

Finally, the EUscreen portal will contain a geographical search functionality, which visualizes search results using the Google Maps API[27]. Only assets that are enriched with geographical information can make use of this search option. Users can enter search queries and use filters to limit the amount of search results returned.

7.2.8 EUscreen and Europeana

EUscreen will provide the technical solutions needed to achieve interoperability between the TV archive collections of Europe and with Europeana in particular. It will collect, harmonise, enrich and make accessible more than 30,000 digital items and will thus have an enormous impact in providing access to the European television heritage.

7.2.9 Handling of Intellectual Property Rights Issues

Television material is most probably the hardest to publish online. This is because rights for

[26] http://www.joanneum.at

[27] http://code.google.com/apis/maps/

programme material are varied and complex, residing with a range of stakeholders including (but not restricted to) actors, programme participants, performers, writers, directors, composers, production companies and, at times, with the broadcasters themselves. The rights not only vary across different European countries, but also often vary widely in each country for historical and commercial reasons. Whereas single state broadcasters in television's early years may have negotiated directly with individuals or agents and unions representing the range of stakeholders working in television, recent deregulation and expansion of products, services and channels means that independent production companies now not only have a commercial stake in the television industry but also add an additional layer of negotiation and ownership. All these different factors mean that the clearance of rights for television programming can be both costly and time consuming.

Nonetheless, the Video Active project has demonstrated that it is possible to clear the rights on hundreds of hours of content without undue expense. This has been particularly the case with a large range of factual material, covering news, current affairs and documentaries. In many cases it has meant, however, that archive partners have not been able to contribute popular programme forms such as drama, comedy, music (pop, traditional and classical) and sport. EUscreen will also develop models for inclusion of non-factual and entertainment forms of content where holdings and IPR allow.

EUscreen has set up a Working Group that will assess IPR and rights issues in the audiovisual field, especially focusing on the complex circumstances of Internet based use and creative reuse. It will explore, in the context of the rapidly changing media landscape, new developments, opportunities, challenges and solutions to existing rights issues. It will take into account the point of view of all actors in the media landscape, and develop strategies, recommendations, checklists and guidelines for dealing with rights issues in the long-term (such as through best practices, legal arrangements and design solutions).

EUscreen and the European Film Gateway will hold a combined workshop on copyright issues and clearance in October 2010.

7. 2.10 Contacts

Link: **www.EUscreen.eu**

Contact information

Project co-ordinator
Prof. dr. Sonja de Leeuw
J.s.deleeuw AT uu.nl
University of Utrecht

Technical co-ordinator
Johan Oomen, MA
joomen AT beeldengeluid.nl
Netherlands Institute for Sound and Vision

7.3 DISMARC

7.3.1 Introduction

Music audio is aggregated to Europeana via the Europeana Audio Aggregation Platform (AAP), which uses the infrastructure of DISMARC to harvest, check and upload music audio and other related content, and metadata.

7.3.2 Types of content

The selection criteria require that content should correspond to the Commission's definition of Cultural Heritage content. Catalogue data-only collections can be aggregated to DISMARC but Europeana requires the existence of underlying audio as well as catalogue data, thus catalogue data-only collections will not be searchable via Europeana.

The AAP aggregates digital music audio and related content, including album covers, photographs, artist biographies and discographies. See below for formats.

7.3.3 Network and partner strategy

Content owners provide their content to DISMARC via the DISMARC Starter Kit, and will in the future also be required to sign the Europeana Licensing Agreement (when it is finalised) as a pre-condition for aggregation to Europeana. Details of the Europeana requirements can be found here.

7.3.4 AAP data format

All formats of **metadata** can be accepted – DISMARC will make a data-mapping specific to each content owner. Mapping of content to the Europeana profile (ESE) is carried out automatically.

i. Content ingestion architecture, workflow and tools

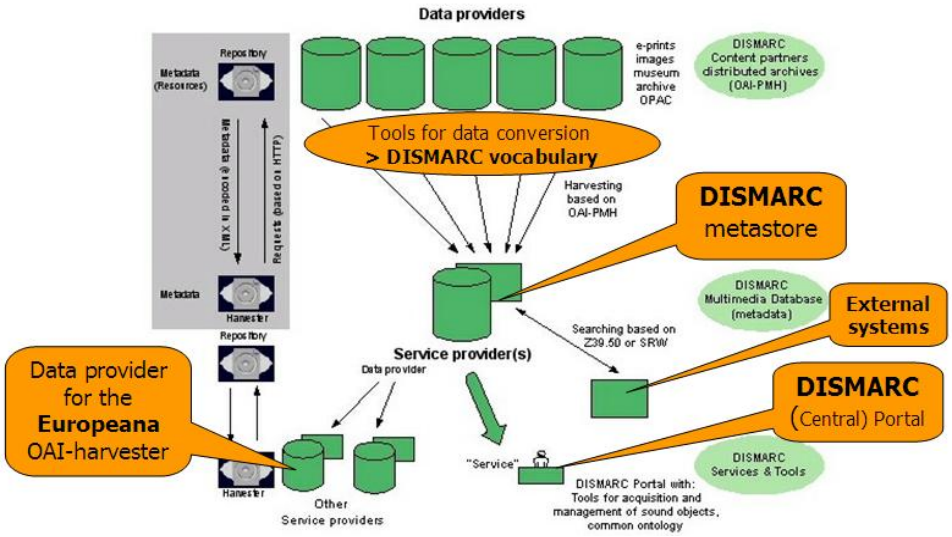

Figure17: Workflow (for the initial ingest)

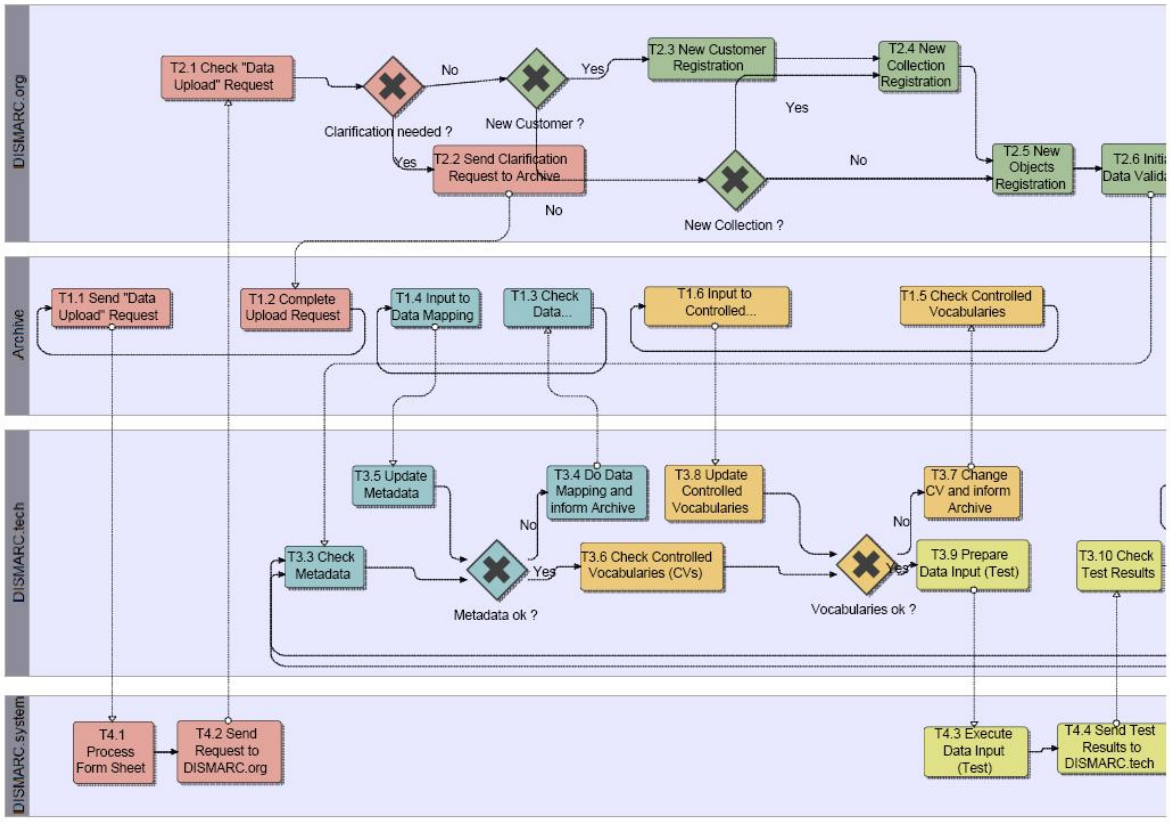

Figure 18a: Workflow

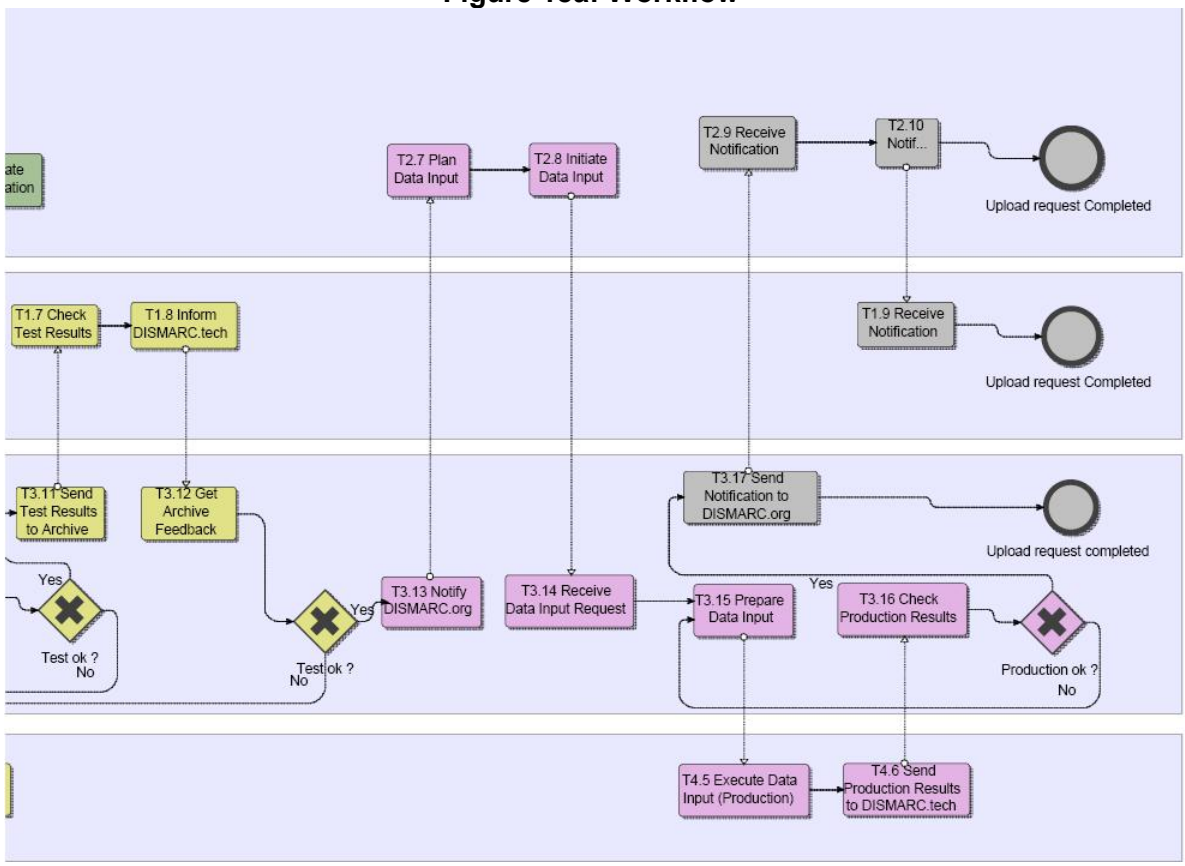

Figure 18b: Workflow (continued)

Future ingestions are initiated by the content providers/archives and communicate directly with the DISMARC system using the established import routines and mappings of this first ingest. Archives have the possibility to install their own AAP OAI system (‘Highway to Europeana on a Stick’).

7.3.5 Language resources used

The DISMARC AAP portal provides the following 30 portal languages:

> Arabic, Bulgarian, Chinese, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hindi, Hungarian, Italian, Japanese, Korean, La TV ian, Maltese, Norwegian, Polish, Portuguese, Romanian, Romany, Russian, Slovenian, Spanish, Swedish, Ukrainian.

The DISMARC metastore (the data repository) provides for multilingual searches in the following 36 languages:

> Arabic, Bulgarian, Chinese, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hindi, Hungarian, Italian, Japanese, Korean, Kurdish, La TV ian, Lithuanian, Luxembourgish, Maltese, Norwegian, Polish, Portuguese, Romanian, Romany, Russian, Slovak, Slovenian, Spanish, Swedish, Turkish, Ukrainian.

Multilingual aspects are provided at different levels –

- users can expand a multilingual word list which contains preferred terms used in partner archives and in the audio domain;
- multilingual vocabularies (e.g. IconClass) can be imported during the input of raw data;
- mapping processes of raw data provided by an archives can be expanded by adding relevant terms taken from a DM controlled vocabulary;
- queries can be expanded by the addition to the search results of terms (via DM controlled vocabularies) in selected language;
- result records show all equivalents of a term in translated languages (provided the terms is included in a DM CV
- the portal language itself can be selected from 20+ languages.

7.3.6 Online accessibility of content – AAP portal specifications

Users can access the metastore via a standard Web browser. An authentication procedure enables individualised access.

The metastore harvests the metadata from any different automated systems (providers) and provides powerful search capabilities for searching this diverse data. This eliminates the need to know how to search objects in individual systems. New metadata is harvested automatically in pre-defined intervals (daily). DISMARC uses the OAI Protocol for metadata harvesting (OAI-PMH).

7.3.7 AAP and Europeana

As noted, the AAP provides audio to Europeana. All aggregation and checking of content is initially done in the AAP, is then checked via the Europeana Content Checker and then provided to Europeana. The AAP provides tools for content providers that allow them to update their content on a daily basis; a procedure for continuous re-harvest by Europeana is under development.

7.3.8 Other guidelines and tools (IPR, digitisation, normalisation and enrichment guidelines, etc)

A extensive guide to the AAP for content owners can be found here. The AAP provides a Dublin Core metadata application profile that is specialised for audio, the mapping of the data to ESE is done automatically. New archives can be given the rights to map their data to the metadata schema via an online Mapping tool. The platform offers an IPR basket that sends out requests for IPR clearance of individual audio files directly to the content providing archive. In addition to that the platform offers the creation of individual licences for pre-defined uses (to date available in English and German).

7.3.9 Contacts

http://www.dismarc.org/info/index.html
Project coordinator: Rundfunk Berlin-Brandenburg, DISMARC

Primary contact

Name and Surname	Martin Gordon (or Johannes Theurer)
Job title	Project manager
Telephone number	+49-30-97993 35370
Address	Rundfunk Berlin-Brandenburg, Masurenallee 8-14, D-14057 Berlin
E-mail address	martin.gordon@rbb-online.de

Technical contact

Name and Surname	Gerda Koch (or Walter Koch)
Job title	Manager, technical team leader
Telephone number	+43-316 83 53 59
Address	Klosterwiesgasse 32, 8010 Graz, Austria
E-mail address	kochg@ait.co.at

8. Europeana and the PrestoPRIME Competence Centre

The European Digital Library Foundation expects to collaborate with PrestoPRIME in order to raise awareness among content providers about the benefits of opening up access to the audiovisual heritage. Europeana and the collaborating projects that we presented so far are mainly involved with access to cultural heritage. There is a wealth of research, development and information produced in this respect in the form of guidelines and reports in areas such as handling of the complex copyright issues, content selection and contextualisation, data modelling for access and interoperability and multilingual search and retrieval of content. The PrestoPRIME Competence Centre could help raise awareness about these to a wider audience of content providers and to provide support and guidance into joining one of the existing aggregators.

Additionally, PrestoPRIME is developing tools and software that can support audiovisual content providers to make their digital contents more widely accessible such as the metadata mapping service from Work Package 4. More broadly speaking, Europeana and the collaborating audiovisual projects can benefit from the work that the Competence Centre is doing with regards to preservation of audiovisual content to ensure long term availability of their contents for access.

It is worth exploring further the synergies between Europeana and the European audiovisual projects representing the audiovisual sector in order to ensure that trusted information and services are produced and shared among content providers across the spectrum of preservation, digitisation and access.

8. Conclusions and future developments

The present document included guidelines for AV content providers on how to make their contents available to Europeana. Europeana's particular technical and organisational requirements were explained but also, since Europeana's content and partner policy largely depends on the collaboration with aggregators, we briefly presented the particular requirements of the major audiovisual content aggregators that currently collaborate with Europeana.

The guidelines reflect the state of affairs as of March 2010 while still a lot of the developments described are in progress. The update of the current document will therefore be necessary in a few months time in order to provide more relevant information to new content providers.

A particular development worth mentioning and which is likely to have an impact on how audiovisual content is accommodated in Europeana, has to do with the implementation of a new Data Model. As it was explained in the document, Europeana's current data model is domain-agnostic. This is likely to change in the future as the Europeana new Data Model evolves to incorporate richer and domain-specific knowledge. The future Europeana Data Model (EDM) intends at bringing much more flexibility to Europeana ingestion and exploitation procedures. It also aims at enabling links across objects, as well as links between objects and explicit representations of contextual entities (events, persons, places) that will form richer semantic networks around them.

In particular, EDM will allow each provider to submit metadata that follow their own practice, as long as the schemas structuring them are mapped to a core of EDM classes and properties. It is also possible that several partners gather and design a specific "profile" for EDM, which will allow them to submit metadata that match the requirements expressed in a given community, for instance one which focuses on audio-visual documents.

In the coming weeks, there are plans to have EDM undergo a series of community-specific reviews (libraries, audio-visual institutes, archives and museums) which will investigate whether there is a conflict between EDM features and the core requirements of given communities. These reviews may ignite the process of creating several of the aforementioned "profiles," one of them could be dedicated to the description of audio-visual material.

9. Annex

Projects' consolidated information table:

	EFG	EUscreen	DISMARC	Europeana
Type of content	Film and film-related material (film clips, pictures, text material)	TV and broadcast archival clips and related material (ie program details)	Audio and music clips	All
Type of content (secondary)		Graphic images, text where related to main content, UGC	Graphic images, text where related to main content	Text, video, audio, images
Metadata format	EFG Format (based on FRBR, EAC, CEN EN 15709).	EBUcore with qualifiers	Dublin Core Qualified (with a few extensions): DISMARC Object Application Profile, DISMARC Collection Application Profile	Currently E.S.E. Specifications v3.2.2
Content ingestion tools	Alternatively OAI Harvesting or online upload through OpenDlib, DRIVER.	Complete content ingestion workflow and tools. Supports upload of excel files and harvesting of XML in any schema	Online data upload tools, DISMARC-on-a-Stick (OAI provider & harvester)	Supports harvesting via OAI-PMH of aggregators' repositories
IPR	IPR remains with original IPR owners; rights have to be cleared for use in EFG.	IPR remains with original IPR owners; rights have to be cleared for use in EUscreen.	IPR is unaltered, and remains with the original IPR owners.	Data provider's and aggregators agreements currently being drafted. Will be mandatory for submission of contents.
Access	Via the EFG Portal and Europeana.	Via EUscreen, Europeana and small scale test cases for use in education, research and leisure applications	Access is via the Europeana. Content management takes place via the DISMARC portal (www.dismarc.eu).	Europeana.eu is a generic cross-domain portal.
Storage	Content remains on content providers' own servers. Only metadata and thumbnails stored centrally.	Metadata are stored centrally for the harmonisation, indexing and enrichment. Videos are	Audio can be stored either on the content provider's own server or on the DISMARC Audio server if the content owner has	Only metadata and cached thumbnails are stored, indexed and enriched

		transcoded in standard web format. During transcoding screenshots are extracted from media files to be used for the display of results.	no server. This service is free.	
Controlled vocabularies	EFG multilingual mini-vocabularies for certain types of information such as credit&cast functions, formats etc. International standards (ISO) applied where possible (countries, languages…).	Video Active controlled vocabularies for the metadata elements (Genres, Keywords and Location)	DISMARC controlled vocabularies (Genre, Formats, Eras, Geography, Languages, Cultural Groups etc…) IBSS, ET	Gathering information from partners across Europe for the development of a central pool of vocabularies, and other language resources to enable linked data and true multilingual search and retrieval in the future