



Recherche / Archives :
numériser les images, et après ?

18 / 19 / 20 novembre 09

Research/Archives:
Digitizing images, then what?

INDEXATION DE MASSE ET NOUVEAUX OUTILS DE REPRESENTATION

The latest achievements of the EFG project

Julia WELTER,

Deutsches Filminstitut - EFG.

Thank you very much for the invitation to talk, and my apologies for speaking in English rather than in French, but I think it's better for all of us, for me and for you if I don't try my French.

Before I start speaking a bit more in detail about the EFG project, I would like to mention a project which actually can be seen as the predecessor to EFG, This is the MIDAS project, which was carried out from January 2006 until January 2009, The MIDAS project initiated a union catalogue for film holdings of European film archives called "filmarchives online". It is available under www.filmarchives-online.eu.

The main effort in MIDAS was to bring together data from 18 heterogeneous film archival databases and make them accessible via a single access point. Since the start of the MIDAS project, the Internet as well as the archives have developed further. A higher number of people have faster Internet connections, which allow easier presentation of multimedia content, and offers new possibilities regarding the interaction with the users. So, today Internet users are not satisfied anymore with finding catalogue information only, but they expect to find digitised material and content directly available on the Internet for their research purposes, and prefer the cost-free if possible.

Also - on the archival side - larger digitisation projects have been initiated in the last few years, which when the rights are cleared, allow the archives to make a greater number of digitised materials available via the Internet now. Also, where in the past there were many individual national projects giving access to repositories, the number of projects actually bringing together the content collected in these smaller projects rises. Especially noteworthy here of course is EUROPEANA (www.europeana.eu) which gives access to Europe's cultural heritage available in libraries, museums and archives, including audio/visual archives. So basically they are bringing together archival material across the domains.

It was a logical next step for film archives and cinematheques to also give access to digitised archival material they hold in their collections. So this is tackled in the EFG project actually, which just entered its second project year, being funded for three years altogether by the eContent Plus Programme of the European Union.

So, in contrast to the MIDAS project, the EFG project aims at giving users worldwide the possibility to research and view digitised archival material from their home via the Internet. Also, where MIDAS was focused on film collections, EFG has a wider scope, including not only moving images but also other materials held by film archives, like text documents, and images or in lower quantity sound files.

So other than MIDAS, EFG aims at building a digital showcase for collections of European film archives and cinémathèques. When it will be launched in autumn 2010, EFG will offer a cost-free access point to these digital collections, which are partly already available on the Internet, but have not been accessible via a single web portal so far.

Basically there are two kinds of information which are being aggregated in this project: these are authoritative data for films works and persons, as well as object descriptions and location of objects. The digital objects themselves will not be stored in EFG but only linked to, remaining in their original context.

EFG is not only establishing the web portal for the film archives to present their content, but it is also an aggregator project of EUROPEANA at the same time. Due to the high number of contributing institutions to EUROPEANA, it has proven very difficult to homogenise and make interoperable the cataloguing information that comes from the different domains and institutions. Also it is not scalable to integrate metadata on an institution by institution basis, hence domain-specific aggregators like EFG do some work beforehand in terms of homogenising and cleaning data. You could say that in EFG film archives are cleaning up their own back yard before they deliver data to EUROPEANA. Another reason why aggregators make sense is that, with a portal such as EUROPEANA, you inevitably reduce information to some degree, and to circumvent this we are going to build a portal specifically for film archives and cinémathèques, so that if a user finds something in EUROPEANA from the film archives, he can go to the EFG website for more information, and then again get to the individual side of the archive for even more detailed information.

Filmographic information in EFG is collected from national databases, like for example from the [Ciné-Ressources](#) website of the Cinémathèque Française, or from [Filmportal.de](#), which is the National German Filmography maintained by the German Film Institute. The Danish Film Institute has its own [online database](#) as well as [Cinecittà Luce's database](#). These are just some samples from which databases our partners contribute to EFG. So like I said, these databases can already be found on the web, but they will be made searchable via the EFG web portal, through a single access point to federate digital collections and repositories. In the framework of EFG we are consolidating authority files for film works and persons to be able to provide highly reliable filmographic information coming from Europe's film heritage institutions.

The EFG consortium consists of 21 partners, including 15 archives from 15 countries. The Aational Audiovisual Archive from Finland just joined beginning of September for the second project year. 6 other partners are responsible for technical or organisational support. Also part of the consortium is the EDL foundation, which is behind EUROPEANA, as well as the Association des Cinémathèques Européennes, the association of European cinémathèques and film archives.

Core issues of the EFG project are on the one side establishing metadata interoperability allowing to homogenise information coming from the diverse local databases, and on the other side providing guidelines for rights clearance, as this is especially important with regard to the fact that most film archives do not hold copyright to many of the materials kept in their collections. So the first project year, which ended in September 2009 was very much concerned with laying the basis for common a EFG database, which allows to collect data being delivered by the project partners from their databases into one EFG database. The EFG database model was established in July this year, and first test ingestions have been carried out, starting August with a first set of test partners. Other archives started delivering first exports in October. We are currently in full swing regarding ingesting the delivered data sets into the common EFG database. For this purpose, import filters have to be established that in future will allow an automatic upload of further exports from the respective partners.

EFG is also striving to establish OAI-PMH interfaces that are at the moment not very common in the film archival community, but regarded as the preferred means of data exchange from the local databases to EFG. So most archives will deliver their first data export until January, so that in February EFG can provide a comprehensive set of data from almost all content providers to EUROPEANA. To the users, the collected data including the links to the digitised objects are expected to be available via the prototype version of the EFG web portal in autumn 2010. While our Italian colleagues from

CNR-ISTI are establishing the back-end, meaning the database behind the web portal, a technical subcontractor will start developing the web portal front-end in January, which will actually be the web portal that will be accessible for the users.

Metadata interoperability which is the major concern in EFG, is established by applying the [Cinematographic Works Standard](#). It is a standard that resulted from the standardization initiative of CEN. This standard is supposed to provide a framework for the unambiguous identification of film works that are catalogued under different distribution titles. For example, you have the same feature films catalogued or registered in many different cataloguing systems and many different institutions in many different countries. If you have no proper identifier for them and no proper rules for identifying the original title or for example the German or French distribution title, you get confusing search results and double entries. Hence, a framework for the identification of film works and conversions that are associated with film works is needed. The various manifestations of a film work have to be summarised properly under the same title in order to properly identify it in an aggregated database. In EFG, we therefore consolidate existing filmographic information and link digital resource descriptions to filmographic records. The consolidation process actually requires intellectual decisions, which means that the system automatically recognises person and film work entries that are similar, but the actual decision, if two entries looking alike are actually describing the same person or film must be decided by an archivist, who can merge data records where necessary.

So in EFG we therefore consolidate existing filmographic information, link digital resource description to filmographic records. The consolidation process actually requires intellectual decisions, which means that the system automatically recognizes person and film work entries that are similar, but the actual decision, if two entries looking like are actually describing the same person or film it must be decided by an archivist, who then can also merge the data records if necessary.

The EFG database has quite a deep structure and consists of eight entities. These entities are defined with the help of various standards:

For audiovisual material (moving images):

- AVCreation (defined by EN15 907 CWS)
- AV Manifestation (defined by EN15 907 CWS)

For non-audiovisual material (Text material, photos and sound):

- NonAVCreation (defined by ESE / CWS)
- NonAVManifestation (defined by ESE / CWS)

For persons and corporations :

- Agent (Persons, Corporations) (defined by EAC)

For events (premieres, censorship etc)

- Event (defined by EN15 907 CWS)

Information on archival collections an object belongs to:

- Collection

A further entity which is only very reduced and basically just contain the links to the digital objects :

- Item (defined by ESE)

More information on the EFG metadata model is available publicly on the EFG project website in the "[Guidelines and standards](#)" section. In this context, it is important to state that eventually not all collected data from the archives will be displayed in the EFG portal, but only a reduced set. Still, we are striving to receive as exhaustive data from the partners as possible with respect to other contexts and future projects, where the gathered data might be needed completely.

Very closely linked with the process of data ingestion, is the semantic interoperability within EFG. Semantic interoperability has been established by compiling reduced vocabularies that can be applied in EFG. The main aim of establishing EFG vocabularies is to harmonise the very diverse values used in

the archives' local databases. This means that local values will be matched to an according EFG value before being displayed on the EFG web portal to the public, thus allowing a coherent presentation of terms to the users. Moreover, semantic relationships are being established, which allow to put the various entities of the EFG metadata schema into context.

As mentioned before, another important task in EFG is providing guidelines for copyright clearance to the project partners. Most - not all - archives are facing the problem that they hold copyright only to a very limited set of their collections. In many cases, estimated 80-90% of archival holdings have other copyright owners. With film and film-related material is even more complex as usually multiple people are considered rights holders of a film work. For most European countries the director, the director of photography, the screen writer as well as the composer of the musical score are considered creators of a film work and therefore hold the rights to the film. As to the fact, that copyright usually expires seventy years after the death of the last surviving creator, only very few film works are freely available in the public domain in Europe today. Adding to this the creators are often very hard to locate, and it therefore is very difficult to find out whether a film can be seen as an "orphan" or not.

Also, filmarchives hardly ever do a proper rights documentation as this is very time-consuming and expensive, due to the multitude of rights holders to a film work but also due to copyright is subject to change. So it is very demanding to keep such a database up-to-date.

Unlike in the US, where all films before 1923 are being considered public domain, regulations are stricter in Europe, which doesn't make it clearer with regard to rights clearing.

In order to tackle these problems, we did a survey among the EFG partners about copyright regulations in the European countries concerning the use of public domain works, of copyrighted works and orphan works, especially with regards to film archives. The resulting report is available publicly on the EFG website in the outcomes section. The findings from this survey will provide several general guidelines for film archives, basically about what to do and how to proceed when researching cleaning and negotiating rights towards Internet access, into a kind of checklist: "What do I have to have done before being able to publish material on the World Wide Web?" Preliminary guidelines have been provided in a second report available for the EFG partners. They were also introduced during a workshop on copyright clearance just a few weeks back in October. It is envisaged to have another copyright workshop, with more concrete guidelines later in 2010 together with the EUScreen project (follow-up of the VideoActive project).

A brief overview on the activities carried out so far in the EFG project:

July 2009 → finalised EFG database model

July – September 2009 → first test ingestions of partners' data into the EFG database

Starting October 2009 → further ingestions from more partners

November 2009 → test ingestion of data collected so far in EFG to EUROPEANA web portal

Outlook:

January 2010 → start of the development of the EFG web portal front-end

February + May 2010 → further data contributions from EFG to the EUROPEANA portal

Autumn 2010 → second release of the EFG database models including some adjustments

Ca. October 2010 → launch of EFG web portal

On-going until August 2011 → further data ingestions and consolidating film work and authority files as well as linking as well as establishing relationships between objects and film works and/or person records

This is the project website where you can find more information on the project www.europeanfilmgateway.eu . If you are interested in joining EFG with your archive and you are from a yellow country contact us via efg@deutsches-filminstitut.de Thank you very much for your attention.

Suivi éditorial : Lorraine Pereira – chargée de mission pour le patrimoine cinématographique / INP.