

Workshop Metadata Management in Film Archives: Putting the „Cinematographic Works Standard“ EN 15907 to use and introducing the new FIAF Cataloguing Manual, 29-30 March 2017

## ABSTRACTS

*SESSION 1: EN 15907 in practice: Existing applications of the Cinematographic Works Standard in film archives. Experiences, issues, challenges and next steps*

**Georg Eckes (Deutsches Filminstitut, Frankfurt)**

**My first eight years with the Cinematographic Works Standard. An attempt to recap and assess the genesis and impact of EN 15907**

With the ubiquitous use of databases and collection management systems in film archives, the issue of how to maintain records about film works – including multiple versions of a work – has increasingly become a question for information and technology specialists. However, while the distinction between a work and a print has been implemented fairly consistently in the databases of most film archives, the documentation of “non-original” versions of a film work was implemented in many different ways, if at all. For a long time, moving image cataloguers have been lacking guidance on how different versions and other alterations that can occur during the lifecycle of a cinematographic work should be properly identified and recorded. As a consequence, the information gathered by film archivists cannot easily be shared between our institutions – and often not even within a single institution.

To tackle this issue, archives from several European countries joined forces in 2005 and formed a group of experts which worked on a technical specification that was published in two European standards: EN 15744 (2007) and EN 15907 (2009).

The presentation focuses on the genesis and role of the more extensive EN 15907 standard. It does not provide specific implementation examples from the Deutsches Filminstitut. Rather, it is my objective to recap some of the cornerstones in CWS development and adoption, and to assess its implementation and implementation issues. To this end, the results of two surveys among (mostly European) film archives carried out in 2013 and 2017 will be presented and discussed.

**Stephen McConnachie (British Film Institute, London)**

**Five years of EN 15907 at the BFI: how we learned to stop worrying and love the CWS**

1. Transformative data projects are change management projects: how we implemented EN 15907

Describes the BFI’s journey from a traditional, internal-facing, informal documentation model to a standards-facing, centrally governed, formalized model. Outlines the challenges experienced in shifting the systems and the organization, and some practical ideas for facing them.

## 2. The perfect is the enemy of the good: what we learned from implementing EN 15907

Describes a pragmatic approach to standards-compliance, accepting that real collections data at scale will progress towards standardisation slowly, with obstacles and resistance. Embrace imperfection and don't let it block you from acting. Standardisation is a never-ending process, an institutional behaviour, not a finished state.

## 3. How EN 15907 might be optimized (a BFI perspective)

Highlights some of the major challenges for EN 15907 and film metadata practice, and signposts some paths toward solutions.

### **Elżbieta Wysocka, Joanna Kaliszewska (Filmoteka Narodowa, Warsaw)**

#### **Mapping Filmoteka to digital era. Polish National Film Archive's implementation of „Cinematographic Works Standard“ EN 15907**

The Digital Repository Project, initiated by Filmoteka Narodowa in 2010 and ended in 2015, aimed at digitization, digital archiving, cataloguing and making Polish cinema available to a broader audience. After accomplishing a few digitalization projects, a large volume of data in varying file formats had to be managed. Therefore, following the digitization process, this funds helped to build the MAM software, maintain digital collection and we started integrating the metadata model EN 15907 (“CWS”) in 2011, immediately. I would like to present our implementation of the standard from the perspective of our database interface - MAM.

In my presentation I will mention the most difficult elements of our implementation, which were of interpretative, database programming nature and organizational. In the first group I would point out: mapping “traditional” catalogue to FRBR based concepts; the interpretation of unusual materials; vocabularies – became a blend of old and new ideas and often differ from suggested in FIAF Cataloguing Manual (2016).

Problems of a programming nature were: complex metadata structures resulting in complications in searching for information, as relationships between objects occurs not only in the hierarchical relation but in any kind – which caused our database to use graph structures for semantic queries.

Organizational difficulties are: training of employees after change in the rules of inventory, and extending their workflow placing greater demands on their digital literacy and finding consensus in the interpretation of standards (in a large institution with many departments).

Despite major difficulties, thanks to an improved data structure and new tools, the previous separation of objects and activities is developing into one coherent workflow, with a more clearly expressed double nature of our collection (analogue and digital) as well decision making process, parameters used, time taken and people involved.

**Annette Groschke (Deutsche Kinemathek – Museum für Film und Fernsehen, Berlin)**  
**Overcoming challenges of the 3-Level Hierarchy Model of En 15907 at Deutsche Kinemathek**

Our goal since implementing EN 15907 was to use a 3-Level Hierarchy Model of the standard. Once we started entering information in our new Adlib database we encountered challenges when trying to record crossover data regarding the variants embodied on manifestations. Furthermore we had to decide on a solution for documenting compilations that would result in a clear and comprehensible structure. While trying to reach maximum usability we realized the restrictions of the 3-Level model. To overcome our challenges we consulted the FIAF cataloguing manual. Our presentation will summarize the three different approaches we tried and our final conclusion how to proceed in the future.

**Laurent Bismuth (Centre National du Cinema et de l'Image Animée, Bois d'Arcy)**  
**The Variant entity: a possible landmark for a better knowledge of the film collections**

CNC has been aboard the Cinematographic Works Standards (CWS) Project since the very beginning. Alongside BNF and INA, our institution represented France within CEN TC 372, responsible for achieving both EN 15744 and EN 15907, and their Dissemination Activity Programme - when the “Promoting the use of European Standards on the Interoperability of Film Databases” project Team (PT) was launched in 2010, CNC provided one of the 4 hosting experts, and organized a Workshop in Paris that was acknowledged as a success.

CNC's genuine commitment to the CWS is the result of our belief that the standards present solutions to many issues that Film Archives and Cinémathèques confront on a daily basis. One of our major problems is having to deal with different versions - short, long, censored, and so forth - of the same film, since it frequently occurs in our collections. The problem happens to be multiplied when it comes to restored versions, whether we simply hold them or produce them, as it has been more and more the case for several decades now. And when the same film has been restored several times in 40 years, the difficulties can be countless. Using screenshots of our current database, we will explain the limits we meet in identifying and handling those different versions by using mainly a carrier-oriented approach. We will use the records of two films as examples: *Le Joli Mai*, that concentrates length, censorship and authorship issues, and *L'Inhumaine*, which has given rise to many restorations already, in-house or by other operators, that are as many milestones in 40 years of the history of our institution. We intend to demonstrate how the Variant entity constitutes precisely the answer to those issues, by presenting what the situation of both films becomes in the new database that CNC is currently developing with Cinémathèque française, and that implements all of EN 15907 entities.

Although we were – and still are – convinced that CWS's four levels for description match the life cycle of the film, before and after it enters the collections of a Film Archive, we encountered many difficulties while adapting this architecture to fit a complex computer model. Screenshots of successive deliveries of our new system (still in progress) will show the price that we paid for both our level of requirement and our dedication.

In order to share our experience wholeheartedly with the workshop speakers and attendees, we will testify how in that process, we felt that the Manifestation entity was the most uneasy to apprehend, maybe because our current database already embeds this concept, although in an implicit – or unspoken – way. In this regard, the very point of our presentation would be to explain how this long journey has convinced us that, especially when confronted to the very materiality of the film collections, the Variant would be the main added value of the CWS, since it is the entity that best allows us to render the life of both the Cinematographic Work and the institution that holds it.

### **Thelma Ross (Museum of Modern Art, New York)**

#### **The FIAF Moving Image Cataloguing Manual: Your golden ticket to better cataloguing and metadata**

The FIAF Moving Image Cataloguing Manual is the natural companion to the CWS, and is recommended reading for institutions that have adopted or are considering adoption of the metadata schema standard (and even those who haven't or won't!). The presentation will provide details about benefits of the Manual, including:

1. Offers a common language and framework for us to talk and think about our moving images, as well as the foundation for a globally shared and consistent approach to cataloguing
2. Is nearly a one-to-one mapping to the CWS
3. Loaded with definitions, examples, and reasoned recommendations while remaining flexible and sensitive to institutional context
4. Addresses analogue and digital moving images
5. Guarantees intellectual and physical control over your collections
6. Specifies linking to moving image-related documentation and collections
7. Legitimizes arguments for standardizing your moving image metadata

The Manual also has its limits; some areas were considered out of scope and excluded whereas some existing sections may benefit from revision and expansion. Examples include: partial value lists (where LOD may come to the rescue); shot-level description (the promise of automated metadata extraction and standardization?); "Content" entity (where existing but incomplete form/genre/subject heading standards/lists are pointed to); and, digital preservation metadata and/or the relationship between your CMS where the cataloged data lives, and your MAM/DAM where your technical digital data lives.

### **Rubén Domínguez-Delgado, María-Ángeles López-Hernández (University of Seville)**

#### **The attention to the film content on EN 15907 and the FIAF Moving Image Cataloguing Manual**

In the field of library and information science, content analysis is a fundamental task for the efficient retrieval of information by the users of information systems. The Cinematographic Works Standard EN 15907, as well as the new FIAF Cataloguing Manual have demonstrated to be really useful for film

archives in the process of film cataloguing, although they are still being implemented in most of them and much work still needs to be done. However, beyond the technical or formal aspects of films contemplated on these standards, there are some important aspects about the film content analysis which have not been had into consideration.

Due to the insufficient attention to the film content on these standards, users of most film archives cannot currently retrieve on their databases shots, scenes or sequences of movies about a concrete need. In Spain and many other countries, a user who visit today an important TV archive can retrieve on its databases a concrete shot of any personality, place, time or issue. However, this type of retrieval of film information is not possible at most Spanish film archives.

At the Film Archive of our Faculty of Communication Studies (University of Seville, Spain) we are starting to apply the new film standards from some months ago, adding three new fields to our traditional film cataloguing records at the moment. However, paying attention to the needs of our users, and as these new film standards do not contemplate a rigorous film content analysis, we have decided to create a “themed film dictionary”. This is a tool for our users to find lists of film titles according to the issue or theme they need to research about it. Although the results have been very positive, we really feel that this is not enough to cover the needs of the 21st century film moving images users and researchers.

That is why we would like to share our experiences - at the archive and as researchers and professors in the field of Moving Images Archives - and propose a discussion, in the context of this FIAF workshop, about the possibilities of adoption of more film content fields on the film cataloguing standards, what would suppose better possibilities of retrieval of film information by the users of film archives, adapting to their needs and to the new era. The development of new technologies, which are already used at TV archives, could help us for this purpose.

## **Pablo Uceda Gomez (Europeana Association, The Hague)**

### **Data quality, enrichments and entities**

Although difficult to define, data quality is an essential factor to build a powerful database. Europeana can enrich the provided data and add richer information to the records but to do so the data need to be provided in a certain way. In this presentation I will explain how the enrichment system works and why this is important for us and for you.

*SESSION 2: No archive left behind? – The situation in and needs of film archives that have not (yet) applied the CWS*

**Ignacio Lahoz Rodrigo (CULTURARTS – IVAC, Valencia)**

### **Wish CulturArts-IVAC implemented CWS**

In CulturArts-IVAC –the Valencian film archive- we are still working with different document – oriented databases. The film collection, the library and the non-film collection are being catalogued on their own, isolated application. They implement different standards or, as it is for film collections, any of them and have different capabilities for the internet. Consequence of this unstructured situation are the many obstacles that make difficult to share our cataloguing work (internal and externally) and to offer to the users a whole and coherent information of our collected heritage. Added to these inefficiencies, the “system” is very expensive because all of these applications are commercial software that we have to pay every year, per records or per user licences.

The lack of standards and interoperability also means many difficulties when you want to collaborate in shared projects. To export our data we need to map our metadata to another database structure. The problem emerges when the partners don't have a standard to organize it. Then, the discussion on the metadata becomes a time-consuming task to eventually create a new database for an specific purpose that suffer from the same limitations than those that so hardly have transferred its information. And the mapping that each partner has had to do is so on purpose that it will not be useful for another project. Once more, working without common tools results in the waste of efforts and economical resources.

The solution, and also the challenge, is to implement the CWS and take profit of all the advantages that the standards and the interoperability implies. As the CWS metadata are thought to describe all the different classes of items held belonging to a work, it will simplify and improve the task of both film and non-film collections cataloguers in CulturArts-IVAC. In addition, the FIAF M.I. Cataloguing Manual is an excellent guide for the daily work. Sharing and research could be now easier and more complete and reliable.

On the other hand, the CWS definition for the databases interoperability places us again in a more efficient position to carry on collaborative projects. If the Spanish film archives had had implemented CWS, the management of the inventory of the heritage we preserve would have been quicker and complete. By reducing the effort of each partner in preparing the information export we could have focused before on another essential challenge: how to share and access our work and how make it available for citizens.

The lack of money, the lack of specialized personnel and the extreme effects of the economical crisis on the Spanish film archives, are the main reasons we are not using the CWS yet. Furthermore, there is a greater need and sensitivity to the implantation of CWS for the film cataloguing than in the library or the non-film collections, given the huge historical difference between the development of resources and commercial products for libraries and for film cataloguing and management.

CWS is not only the core of the collections cataloguing and access. Around them it is possible to organize a whole system that integrates many of the management tasks that a film archive has to do.

From loans to reproductions and migrations, screenings, use of film footage, rights and acquisition details...

Finally, CWS is an essential tool for the safeguarding of the European Film Heritage. Citizens access to their film heritage is what justifies the film archives' mission and access depends on a coherent and useful cataloguing. For facing the internet challenge on the access to the European films, through several recommendations and resolutions EU has drawn a path in which CWS is one of the keys: gather and describe the films and build a net of interoperable databases that includes the whole European Film Heritage. As CulturArts-IVAC preserves a humble part of this Heritage, we need CWS.

### **Sarah Clothier (American Film Institute, Los Angeles)**

#### **Metadata Challenges and Agent Disambiguation in Re-Architecting the American Film Institute's Catalog of Feature Films**

The American Film Institute's AFI Catalog of Feature Films: The First 100 Years (1893–1993) is an authoritative database of American motion pictures, with over 55,000 records featuring credits, plot synopses and extensive, academically researched production histories. The AFI Catalog was established in 1968 as a response to the U.S. government's mandate to preserve the history and culture of American film. First published as printed encyclopedic volumes, the AFI Catalog was later migrated to an online relational Microsoft SQL database in 2003. AFI is now rebuilding the AFI Catalog on Microsoft's Azure SQL Server Platform-as-a-Service cloud to present a rich media experience, with a robust API that can ultimately allow access to AFI's archival media assets and library collections. To create an infrastructure relevant for the future and implement best practices developed by a global community of film archive experts, AFI is exploring the Cinematographic Works Standard EN 15907 (CWS).

AFI is performing a deep analysis of the AFI Catalog data, identifying metadata standards for its media and archival resources. The existing metadata, which exists purely as SQL table fields, is inadequate when the goal is to create Linked Data and improve disambiguation. AFI intends to utilize CWS as the underlying Descriptive Metadata structure for its existing and re-architected data and metadata. Other standards such as the International Standard Audiovisual Number (ISAN) and the Entertainment Identifier Registry (EIDR) provide compelling frameworks for consideration as additional standardization and Linked Data.

Implementing standardized metadata for Film Titles to link to third party data sets appears to be relatively straightforward. However, the metadata associated with names for Agents such as People and Corporate Bodies can be complex. For example, Richard Burton appears in the AFI Catalog as an actor, a still photographer and a sound editor. Attributions for a widely known Agent such as Burton may be easy to categorize, but for the below-the-line "John Smiths" of the world, accurately crediting their contributions can be challenging. Agent disambiguation is also vital to other aspects of archival work, such as determining copyright claimants. The disambiguation of duplicate Agents for personal names appears to be a universal challenge for the motion picture industry, as well as for film archives. EIDR utilizes a hub and spoke architecture for Linked Data, specifically for Movie Title disambiguation, and the International Standard Name Identifier (ISNI) utilizes a similar architecture

for names. Building upon these two successful models by sharing information among sister institutions such as the British Film Institute (BFI), MovieLabs and ISNI, to name a few interested parties, we may ultimately resolve this problem.

*SESSION 3: New ways of managing, storing and sharing metadata in film archives: Linked Data, RDF and concepts beyond the Relational Database paradigm*

**Detlev Balzer (Deutsches Filminstitut / Independent programmer, Lübeck)**

### **Going global. Reciprocal enrichment of archival descriptions through Linked Open Data**

Modern metadata specifications such as EN 15907 define numerous elements that can be used as connectors to entities from other knowledge domains. An obvious example is the sharing of so-called authority data with other institutions such as libraries. A considerable volume of high-quality authority data has become available as Linked Open Data (LOD) for everyone to use and to enrich.

Using LOD resources can not only reduce the burden of maintaining in-house authority data, but also become a reciprocal process in which new information is added to the global knowledge graph. A case in point is the Gemeinsame Normdatei (GND), a vast LOD resource maintained by the German National Library. Over the past few years, the Deutsches Filminstitut has integrated co-maintenance of the GND into their film cataloguing process.

Beyond institutional cooperation, the LOD realm now offers plenty of opportunity for more informal giving and taking. A few of such opportunities of possible relevance to film archives will be illustrated in this presentation.

**Bram Biesbrouck (Cinémathèque royal de Belgique, Brussels)**

### **Reinventing film archives: linked data, artificial intelligence and robots**

I will present, demonstrate and discuss a roadmap and a futuristic vision on film archives in three chapters.

First of all, an example on how to take advantage of linked data on a small scale and in a disconnected manner. When you're small and simple, you need to abuse data in every way you can to become competitive and elevate your collection to the next level. I'll show you how. On a second - more political - level, I'll explain how to build bridges using linked data by automatically interlinking collections and start cooperating on a global scale. We'll talk about collection interlinks, reasoning and who's the boss. Lastly, a futuristic technology demo and some very good news: starting next year, robots will take over the boring parts of your job. Or: the present state of the art in automated, linked and intelligent film metadata extraction.

This session is for non-technophiles who want to know the benefits of linked data, how to exploit them and what the world will look like tomorrow.

**Harald Sack (Karlsruher Institut für Technologie, Karlsruhe)**

**Everything you always wanted to find \* but were afraid to search**

Popular Video on Demand platforms are all facing a similar problem: how to provide the users with potentially interesting movies that fit to the users' (current) interests and preferences. Moreover, how enable the users also to achieve an overview of all the available archive content without to overstrain or being boring. Guiding the users through the archive along (previously unknown) paths reflecting the user's interests and putting the presented documents in a content-related perspective cannot be obtained by traditional retrieval, but also requires guided browsing and intelligent recommendations, leading to a so-called exploratory search.

To enable exploratory search in video archives, document content has to be annotated with descriptive and machine understandable metadata. Since manual human effort is limited and costly, automated state-of-the art video and audio analysis methods are applied to obtain descriptive metadata of heterogeneous quality, accuracy, as well as reliability and confidence. For successful metadata integration, semantic analysis relates the (raw) media analysis metadata to knowledge bases and encodes the annotation as reusable and machine understandable data that can be exploited for semantic search, content-based recommender systems, as well as exploratory search.

This presentation will provide a brief overview on state-of-the art technologies of video analysis and subsequent semantic analysis for the automated generation of enriched Linked Data based metadata. Furthermore examples are presented for semantic search systems, content-based recommender systems, and exploratory search scenarios.