



Metadata Management for Film Archives Abstract  
Session 2 or 3

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

Scott Bly: Director, AFI Information Technology, [sbly@AFI.com](mailto:sbly@AFI.com)  
Sarah Clothier: Interim Editor, *AFI Catalog of Feature Films*,  
[sclothier@AFI.com](mailto:sclothier@AFI.com)

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).

***What are your plans concerning metadata management and EN 15907 in the future? In which ways can resources be pooled to achieve progress?***

AFI is performing a deep analysis of the *AFI Catalog* data, identifying metadata standards for its media and archival resources. The existing

American Film Institute  
2021 North Western Avenue  
Los Angeles, CA 90027  
323.856.7600 Phone  
[AFI.com](http://AFI.com)

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.

***What are the challenges and possibilities for managing and sharing of film-related metadata? What is your vision of future cinematographic metadata aggregation? Do you have concrete plans to go Linked (Open) 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.