

ELECTIVE COURSE PAPER

Using IFLA's General Principles to Evaluate Moving Image Metadata Schemas

IS 289: Media Description and Access

May Haduong

December 2020

The process of cataloging moving images can take many different forms, largely depending on the metadata standards being used. Standards for the data's structure, value, content, and encoded data format can vary widely, and the differences between them can result in contrasting levels of effective description and access. An analytical look into the standards can reveal the advantages and disadvantages for the collection, its catalogers, and most importantly, its users. This paper will look at metadata structure standards in particular, through a critical analysis of common metadata schemas for cataloging moving images—MARC, Dublin Core, PBCore, and EN 15907—using the IFLA's general principles as criteria to expose the varying levels of the standard's quality and effectiveness for description and access.

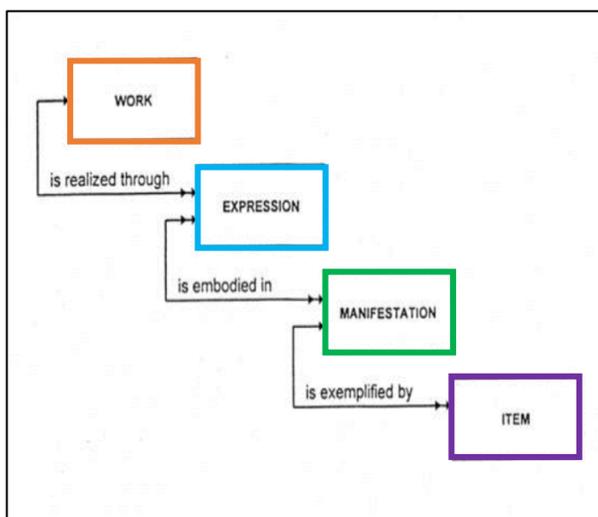
Data structure standards, also described as metadata schemas and element sets, control the organization of metadata by functioning as “containers” of the metadata information.¹ The choice to implement one schema over another usually depends on the materials being described and the nature of intended access. For this paper, the decision to look into these four metadata schemas is based on the different features of each standard and the desire for variation in their origins and purposes; two schemas are designed for wide application (MARC and Dublin Core), whereas the other two are

¹ Anne Gilliland, “Setting the Stage.” In *Introduction to Metadata*, edited by Murtha Baca, 3rd ed. Los Angeles: Getty Publications, 2016, <http://www.getty.edu/publications/intrometadata/setting-the-stage/>.

intended for audiovisual materials specifically (PBCore and EN 15907). The structure of each schema is distinct, based on the intended purposes and concepts that formed its creation.

MARC, known as the machine-readable catalog, is one of the most widely seen metadata standards in and across collection catalogs. The use of MARC can be complex, although its purpose is simple. It is intended to allow for bibliographic, authority, and holdings records to be read by a machine and consequently become intelligible to users. In many ways, MARC supports the Functional Requirements for Bibliographic Records (FRBR) conceptual data model (Figure 1), meaning that it allows for a work, expression, manifestation, and item to be expressed in a single record (see Table 1). Two key aspects of MARC most relevant to this analysis are, first, that MARC is not designed for a specific format but instead can be used broadly regardless of the information object's carrier; and second, it is primarily aimed for machine comprehension and then for human comprehension.

Figure 1. FRBR Group 1 Entities and Relationships



IFLA Study Group on the Functional Requirements of Bibliographic Records, "Functional Requirements of Bibliographic Records: Final Report," *International Federation of Library Associations and Institutions*, September 1997, <http://www.ifla.org/VII/s13/frbr/frbr.pdf>.

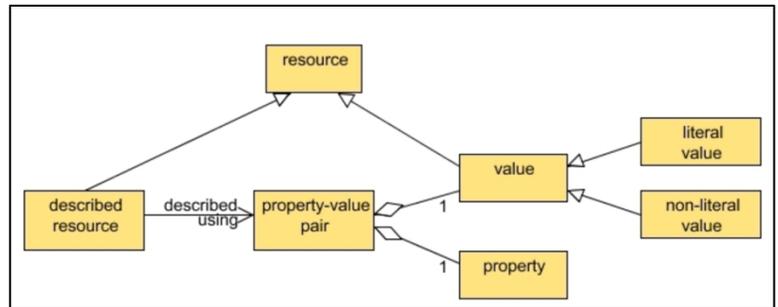
Table 1. FRBR Group 1 Entities Expressed in MARC Records

MARC Field	FRBR Group 1 Entity
1XX	work, expression
240	work, expression
245 - 260	manifestation
300	expression, manifestation
3XX	work, expression, manifestation
490	manifestation
5XX	work, expression, manifestation, item
700 - 730	work, expression
760 - 787	work, expression, manifestation
8XX	work, expression, manifestation

Table devised using "FRBR: FRBR, RDA, and MARC," *Library of Congress*, September 2012, https://www.loc.gov/catworkshop/RDA%20training%20materials/LC%20RDA%20Training/FRBR_Module%203_FRBR%20%20RDA%20%20MARC/FRBR%20%20RDA%20%20MARC_studentversion

Dublin Core is another metadata schema known for its extensive integration and use. It's proposed capability to describe almost any information object is the basis of its widespread appeal for collections. The element set of

Figure 2. Dublin Core Abstract Resource Model



Andy Powell, Mikael Nilsson, Ambjörn Naeve, Pete Johnston, Tom Baker, "DCMI Abstract Model," *Dublin Core Metadata Initiative*, June 4, 2007, <https://www.dublincore.org/specifications/dublin-core/abstract-model/>.

Dublin Core is grounded on the lowest common denominator principle, attempting to capture the "core" attributes to describe a work. This is established through the fifteen main elements of the structure, all of which are optional and repeatable. The Dublin Core Abstract Resource Model (Figure 2) displays how descriptions are modeled; very simply, the resource is described using a property-value pair, meaning, the property is the aspect being described (eg. title, creator, etc.) and the value is the word(s) assigned to that property to uniquely identify the property (eg. the name of the title, the name of the creator, etc.) The "resource" in the data model could be anything, including the work, expression, manifestation, or item. Beyond the fifteen elements, further details of object description are possible through the use of element attributes, which can establish relationships and context. Two significant features of Dublin Core concerning this analysis are its potential for wide application and its simplicity.

Looking towards metadata schemas created for specific formats, PBCore was deliberately developed for describing audiovisual materials. It is similar to Dublin Core, as it was derived from the standard, but PBCore was more purposely designed to include descriptive and technical fields for audiovisual material description. PBCore was originally created to give public broadcasting organizations the ability to manage the metadata of their media, however, it has been adopted more generally by moving image archives and libraries. It may be

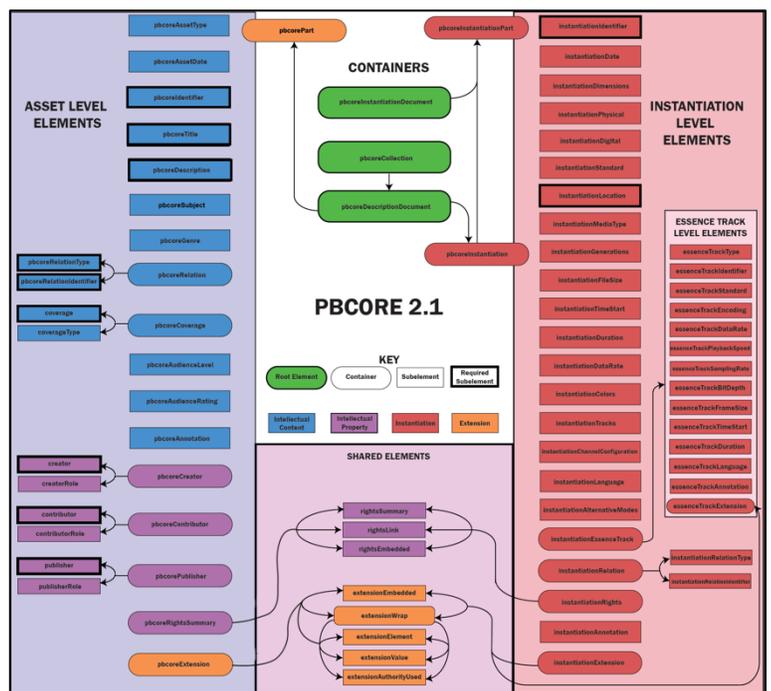
accompanied by other metadata standards for additional needs through extension capabilities, though at the basic level, it is focused on descriptive and technical metadata at the asset and instantiation level, which could also be thought of as the manifestation and item level of the FRBR model (see figure 3).²

The last metadata schema that

will be discussed is the EN 15907,

which is a European standard that structures the description of cinematographic works based on its primary entities, contextual entities, varying element types, and relationships. The primary entities of this standard align well with the FRBR data model, as EN 15907 similarly uses the entities such as cinematographic work, variant,

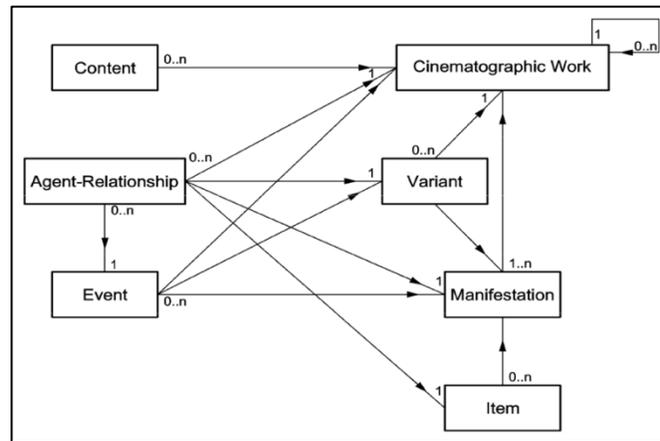
Figure 3. PBCore Data Model



“PBCore Data Model Visualization,” *PBCore*, accessed December 16, 2020, <https://pbcore.org/data-model>

² “Frequently Asked Questions,” *PBCore*, accessed December 16, 2020, <https://pbcore.org/faqs>.

Figure 4. EN 15907 Data Model



"Elements of the Data Model," *filmstandards.org*, accessed December 16, 2020, http://www.filmstandards.org/fsc/index.php?title=EN_15907

manifestation, and item, with the extension of the entities of content, agent, and event to further describe and contextualize. It is useful that on all of these four levels, agent-relationships can be assigned, and for three of the levels, (all excluding the item level) events can be

attached, allowing for a data model more suitable for the nature of moving images. In this way, while the FRBR model can conceptualize description for potentially any information object in the universe, EN 15907's model can only be used for moving image description. Therefore, not only was the schema designed for moving images, but its use is exclusive to them as well.

Evaluating these metadata schemas to expose the varying levels of effectiveness will be done using the International Federation of Library Association's (IFLA) General Principles proposed in the *Statement of International Cataloging Principles* published in 2009.³ The annotated listing of the nine general principles is as follows:

1. Convenience to the user: *How much does the schema consider the user's experience?*

IFLA states this principle is to be most prioritized, as providing ease of use will likely be the most impactful for the user's experience with the catalog. It relates to other cataloging and metadata guidelines, such as those stated in Gilliland's work,⁴ but also supports the user tasks that FRBR defines as

³ "Statement of International Cataloging Principles," *The International Federation of Library Associations and Institutions*, 2009, https://www.ifla.org/files/assets/cataloguing/icp/icp_2009-en.pdf.

⁴ Anne Gilliland, "Setting the Stage."

essential abilities in a catalog—finding, identifying, selecting, and obtaining resources.⁵ Convenience to the user during these research tasks is an essential goal of proper cataloging.

2. Common Usage: *Does the schema prioritize common knowledge and forms of use?*

This principle also relates to Gilliland’s work, when they state, “metadata presents some of the information that an information professional might have provided in a traditional, in-person reference or research setting.”⁶ The schema should outline metadata that is typically of interest to researchers. However, while this principle may be intended to support the broadest range of users, it can also leave room for faulty assumptions about who the users are, what they are interested in, and what “common usage” looks like, especially for a diverse user base. For that reason, this second principle should be considered carefully from many different perspectives.

3. Representation: *Does the schema allow for the resource to be described based on how it originally appears on the resource itself?*

Representing a resource in the catalog is crucial for its discovery and recognizability. Therefore, this principle aligns with the FRBR promoted user tasks of finding and identifying.

4. Accuracy: *Does the schema require a precise description to achieve accuracy?*

Representation of the resource is only of value if that representation is accurate. Once again drawing upon Gilliland’s work, the ability of metadata

⁵ IFLA Study Group on the Functional Requirements of Bibliographic Records, “Functional Requirements of Bibliographic Records: Final Report,” *International Federation of Library Associations and Institutions*, September 1997, <http://www.ifla.org/VII/s13/frbr/frbr.pdf>.

⁶ Anne Gilliland, “Setting the Stage.”

to “certify the authenticity and degree of completeness of the content”⁷ is essential to accurately represent a resource. Achieving this principle in the catalog could be demonstrated by the schema’s allowance of precise elements, sub-elements, or some other form of faceted details.

5. Sufficiency and necessity: *Does the schema only require indispensable elements?*

This principle is interesting in that it recommends only providing data elements “that are required to fulfill user tasks and are essential to uniquely identify an entity.”⁸ Of course, the inclusion of irrelevant data should be avoided, but also data that is too specific or too broad to the point where it doesn’t assist in the finding, identifying, selecting, or obtaining a resource should be prevented.

6. Significance: *Are the elements required by the schema significant to the description?*

IFLA explains this principle as “data elements should be bibliographically significant.”⁹ This is important for facilitating FRBR’s user tasks and also acts as a counterbalance to the previous principle of accuracy. Data elements should be detailed and precise, but not so specific to the point that they are insignificant to the representation of the resource.

7. Economy: *Does the schema allow for efficiency and clear objectives?*

The benefit of the cataloger is a key consideration in this principle. When there are multiple possibilities, IFLA states, “preference should be given to the way that best furthers overall economy (i.e., the least cost or the simplest approach).”¹⁰ This may lead to benefitting the user as well. One criticism of

⁷ Ibid.

⁸ “Statement of International Cataloguing Principles,” IFLA.

⁹ Ibid.

¹⁰ “Statement of International Cataloguing Principles,” IFLA.

this principle is that it may contradict other IFLA principles that prioritize dedicated time and effort over efficiency (for example, ensuring accuracy, necessity, or consistency will likely not be a simple process, but it still included as a recommended principle because of its importance).

8. Consistency and Standardization: *How does the consistency of the schema affect description and access?*

Although metadata schemas are standards themselves, their consistency will likely vary across their application. For example, some schemas allow for optional and voluntary element fields. This principle can guide analysis of how the level of flexibility of the schema's rules play a role in effective description and access.

9. Integration: *Does the schema rely on a shared set of requirements?*

This principle can relate to the importance of interoperability between metadata schemas, as the adoption of common rules across schemas and repositories can benefit the mapping of their metadata. Integration of a common set of guidelines eventually allows for interoperability that results in further access to materials.

While this set of general principles has its imperfections, like the assumption of common use, the contradictory guidelines, and the absence of any principles regarding the importance of equitable description or culturally inclusive description practices, these nine principles have guided international cataloging codes for decades.¹¹ For the following evaluation of MARC, Dublin Core, PBCore, and EN 15907, these guidelines can expose the advantages and disadvantages of their varied capabilities. It has been

¹¹ Ibid.

decided that this is a suitable criterion to address how these schemas determine success for cataloging moving images because of three primary reasons: 1) In general, the set of principles consider both the user and the cataloger. 2) The principles can be applied to online library catalogs. 3) The principles are general enough to encompass all types of materials, beyond textual works, to include moving image cataloging as well. These last two points were not covered in IFLA's original 1961 Statement of Principles; however, the 2009 revised edition has included them as useful and necessary considerations. As the Statement currently stands, these nine principles should be key concerns when assessing metadata schemas.

MARC

Measuring the MARC schema against the IFLA General Principles reveals its few strengths and its several weaknesses. It seems to perform well in the IFLA terms of representation and integration. Its ability to represent a moving image is based on its provided bibliographic fields, such as the general physical description fields (007, 300) that allow for motion picture, playing time (306), projection characteristics (345), video characteristics (346), and so on.¹² It typically relies on a common set of rules, such as data value and content standards (LCSH, LCGFT, etc.), to promote integration. However, that is about the extent of its strengths for moving image cataloging. Most notably, MARC's downfalls appear to be its lack of accuracy, economy, convenience, consistency, and standardization. This evaluation is based on personal use and observation, as well as a recently conducted interview with UCLA Film and Television Cataloger, Amanda Mack.

¹² "MARC 21 Format for Bibliographic Data," *Library of Congress Network Development and MARC Standards Office*, December 2020, <https://www.loc.gov/marc/bibliographic/>.

For Mack, MARC may get the job done, but there are some clear downfalls, such as its lack of accuracy and economy, in particular.¹³ This is likely a result of MARC not being a moving image specific schema. For example, while the MARC bibliographic 511 field is designated as the “Participant or Performer Note,” the UCLA Film and TV archive has locally repurposed that field to be more of a “Cast Note.” Moreover, the 250 field, known as the “Edition Statement” in MARC, actually functions as the “Version Note” in UCLA’s moving image archive catalog. This absence of adequate terminology for moving image cataloging is just one aspect of inaccuracy. There are also limitations of the fixed fields, as Mack expressed the frustration of not being able to properly code a 4 ¾ video format because it is not provided as an option. So, while the representation of moving images may be possible, the most accurate representation is not optimized in a MARC record. There are also issues regarding its convenience to the user, just based on the high number of encoded elements that require specialized knowledge to comprehend.

Dublin Core

In contrast to MARC, Dublin Core offers a clearer demonstration of the IFLA General Principles in its schema. Convenience to the user, common usage, economy, sufficiency and necessity appear to be advantages of the schema, based on its plain text elements, simplicity, and the effort to primarily contain the core fundamental metadata. It’s simplicity also allows it to be flexible, extensible, and compatible, allowing for a higher degree of integration.¹⁴ The simplicity of Dublin Core has been successful in some of these regards; however, its simplicity can come at a cost. As discussed by Julie

¹³ Amanda Mack (Cataloger, UCLA Film and Television Archive), interview by Jessica Craig, December 7, 2020.

¹⁴ Julie Weagley, Ellen Gelches, and Jung-Ran Park, “Interoperability and Metadata Quality in Digital Video Repositories: A Study of Dublin Core,” *Journal of Library Metadata* 10, no. 1 (2010): 37–57.

Weagley, Ellen Gelches, and Jung-Ran Park, Dublin Core's implementation across repositories can reveal some varying levels of completeness, accuracy, consistency, use of controlled vocabularies, and interoperability. For example, in regard to the completeness, they state: "Of the fifteen elements suggested by Dublin Core, only Title is supplied across the repositories 100% of the time. Following Title are Description (99%), Date (96%), Identifier (83%), Type (83%), and Relation (83%) ... Four elements are collected at less than 50%, Creator, Contributor, Source, and Coverage." Not only does this reveal that Dublin Core records often don't use all 15 elements, but the consistency of which elements also differs. Another major issue, mentioned by Weagley, Welches, and Park, is the semantic vagueness of some Dublin Core elements, for example, Format, Type, Contributor, and Creator. This ambiguity can be a major downfall for integration. However, as they conclude, while the Dublin Core schema shows low consistency across repositories, it is actually quite high within them. Thus, the scope of the effectiveness of Dublin Core should be completed with this in mind. Overall, Dublin Core seems to reveal more advantages than MARC, while still not being most capable for moving image cataloging.

PBCore

As mentioned, PBCore was derived from Dublin Core to offer moving image catalogers a more appropriate schema for audiovisual formats. Its specific application promotes its convenience to the user, use of common terms, bibliographical significant elements, and economy. It's encoded in XML on the back end, which allows for more easily shared collection metadata and integration. As Rebecca Fraimow expressed, the XML data format allows for the inclusion of complex technical and descriptive metadata without bombarding the front-end user experience with information they

may not be interested in.¹⁵ However, this effort towards convenience brings up questions of common use. In the American Archive of Public Broadcasting, much of the technical metadata found in the XML file is not all displayed on the front end, since it is assumed to be typically uninteresting (see figure 5). The encoded information still exists, but it is not made readily available to the user. If the inclusion of technical metadata is one of the reasons for PBCore's creation in the first place, perhaps it should be more accessible. The decision to exclude some of the technical metadata may align with the common use principle because perhaps, the technical information truly isn't commonly used. But the decision is backed by assumptions that should be explored. Largely, PBCore is a useful schema for moving image catalogs, as it considers the user, includes accurate moving image terminology, allows for significant elements and optimal representation, and promotes integration.

Figure 5. American Archive Front Plain Text Record

Transcript	Hide	Description
Series	Eyes on the Prize	Filmed interview with Leola Montgomery for Eyes on the prize. Discussion centers on the Brown vs. Board of Education legal case which she and her husband pursued for the benefit of their daughter, as well as a discussing the segregated school system of 1950s Kansas.
Title	Interview with Leola Montgomery	
Producing Organization	Blackside, Inc.	Created 1985-10-26
Contributing Organization	Film and Media Archive, Washington University in St. Louis (St. Louis, Missouri)	Genres Interview
AAPB ID	cpb-aacip/151-gf0ms3kt2s	Media type Moving Image
		Duration 00:11:13
If you have more information about this item than what is given here, we want to know! Contact us, indicating the AAPB ID (cpb-aacip/151-gf0ms3kt2s).		
Credits AAPB Contributor Holdings Citations		

Front-end PBCore record displaying series, title, producing organization, contributing organization, a description, creation date, genre, media type, duration, credits, holdings, citations, in plain-text readable format. Screenshot excerpt from Rebecca Framow's discussion with May Haduong, November 2020.

Figure 5. American Archive XML Record

```

<-instantiationAnnotation annotationTypes="organization">
  Film & Media Archive, Washington University in St. Louis
</instantiationAnnotation>
</pbcCoreInstantiation>
<-pbcCoreInstantiation>
  <instantiationIdentifier source="MAVIS Component Number">636-4</instantiationIdentifier>
  <instantiationIdentifier source="MAVIS Item ID">5891</instantiationIdentifier>
  <instantiationIdentifier source="Appearance Release">Y</instantiationIdentifier>
  <instantiationIdentifier source="Program Number">101</instantiationIdentifier>
  <instantiationPhysical>Audio cassette</instantiationPhysical>
<-instantiationLocation>
  Vault Site: Washington University Film and Media Archive (West Campus); RackNo: CAS.0579
</instantiationLocation>
<instantiationMediaTypes>Moving Image</instantiationMediaTypes>
<-instantiationEssenceTrack>
  <essenceTrackType>audio</essenceTrackType>
  <essenceTrackAnnotation>Screen direction : R</essenceTrackAnnotation>
</instantiationEssenceTrack>
<-instantiationAnnotation annotationTypes="organization">
  Film & Media Archive, Washington University in St. Louis
</instantiationAnnotation>
</pbcCoreInstantiation>
<-pbcCoreInstantiation>
  <instantiationIdentifier source="MAVIS Component Number">636-5</instantiationIdentifier>
  <instantiationIdentifier source="MAVIS Item ID">62100</instantiationIdentifier>
  <instantiationPhysical>16mm film</instantiationPhysical>
  <instantiationStandard>Film</instantiationStandard>
<-instantiationLocation>
  Vault Site: Washington University Film and Media Archive (West Campus); RackNo: EYES.794.11
</instantiationLocation>
<instantiationMediaTypes>Moving Image</instantiationMediaTypes>
<instantiationGenerations>Original</instantiationGenerations>
<instantiationGenerations>Negative</instantiationGenerations>
<instantiationDuration>0:4:10</instantiationDuration>
<instantiationColors>Color</instantiationColors>
<instantiationLanguage>eng</instantiationLanguage>
<-instantiationEssenceTrack>
  <essenceTrackType>Film</essenceTrackType>
</instantiationEssenceTrack>

```

Back-end PBCore record expressed in XML encoding. This record holds more technical metadata not seen on the front-end of the same record. Screenshot excerpt from Rebecca Framow's discussion with May Haduong, November 2020.

¹⁵ Rebecca Framow (Archivist, GBH), in discussion with May Haduong, December 2020.

Lastly, a look at how the IFLA's General Principles function in the EN 15907 metadata schema. There were some specific strengths called out by the basic structure and design. Like PBCore, EN 15907's specific use for moving image cataloging heavily benefits it's the effectiveness of its description and access to the format. In particular, access through interoperability. For example, while EN 15907's simpler counterpart schema 15944 is similar to Dublin Core but for moving images, 15907 is a more comprehensive schema, as Ronny Lowey explains, "designed to provide a pathway towards increased interoperability, both among film databases, and between these and other information systems"¹⁶ Therefore, EN 15907 aligns well with the IFLA General Principle of integration, common use, and consistency. These features have been demonstrated through its adoption for union catalogs, such as the filmarchives-online.edu and europeanfilmgateway.eu.¹⁷ Further research did not prove any disadvantages of the schema, although the limited amount of literature regarding this topic may be a reason for this. Generally, EN 15907 appears to be an optimal metadata schema for moving image cataloging, especially in regard to enhanced integration and interoperability.

Although the practice of cataloging is full of standards and guidelines, the process and result of cataloging moving images can look very different in and across repositories. The IFLA General Principles have become a useful measure to assess cataloging practices in general, as it can reveal the varying levels of effective description

¹⁶ Ronny Lowey, "CEN Standards for Metadata About Cinematographic Works," June 2010, *Duetsches Filminstitut*, <http://filmstandards.org/media/cen-cws-synop-2010-06a.pdf>.

¹⁷ "Metadata Management in Film Archives: Putting the 'Cinematographic Works Standard' EN 15907 to use and introducing the new FIAF Cataloguing Manual," *International Federation of Film Archives*, accessed December 2020, <https://www.fiafnet.org/pages/Training/Metadata-Management-in-Film-Archives.html>.

and access. For the four metadata schemas discussed in this paper, the IFLA principles were seen demonstrated in various ways. For MARC, although it could represent moving images sufficiently, its major criticisms included a lack of accuracy, economy, and convenience. With Dublin Core, its simplicity aided its convenience, common use elements, and economy, but cost accuracy and consistency. PBCore was much more effective all-around, with optimal convenience to the user, accurate terminology, and economy, although its application may bring up questions of common use. And finally, EN 15907, as a more recent development, demonstrates a high level of integration possibilities, with accuracy and representation also exhibited. In conclusion, it seems as though improvements of metadata schemas for moving images progress with time, as each new structure builds off the strengths and weaknesses of a previous one. With the adoption of principles such as IFLA's, it may be most successful, although, implementation will still likely vary depending on its use.

Bibliography

- “FRBR: FRBR, RDA, and MARC.” *Library of Congress Cooperative and Instructional Programs Division*. September 2012.
https://www.loc.gov/catworkshop/RDA%20training%20materials/LC%20RDA%20Training/FRBR_Module%203_FRBR%20&%20RDA%20&%20MARC/FRBR%20%20RDA%20%20MARC_studentversion_20120818.pdf.
- Gilliland, Anne J. “Setting the Stage.” In *Introduction to Metadata*, edited by Murtha Baca. 3rd ed. Los Angeles: Getty Publications, 2016.
<http://www.getty.edu/publications/intrometadata/setting-the-stage/>.
- IFLA Study Group on the Functional Requirements of Bibliographic Records.
“Functional Requirements of Bibliographic Records: Final Report.” *International Federation of Library Associations and Institutions*. September 1997.
<http://www.ifla.org/VII/s13/frbr/frbr.pdf>.
- “MARC 21 Format for Bibliographic Data.” *Library of Congress Network Development and MARC Standards Office*. December 2020.
<https://www.loc.gov/marc/bibliographic/>.
- “PBCore Data Model Visualization.” *PBCore*. Accessed December 16, 2020.
<https://pbcore.org/data-model>.
- “Frequently Asked Questions.” *PBCore*. Accessed December 16, 2020.
<https://pbcore.org/faqs>.
- “Statement of International Cataloguing Principles.” *The International Federation of Library Associations and Institutions*. 2009.
https://www.ifla.org/files/assets/cataloguing/icp/icp_2009-en.pdf.
- Weagley, Julie, Ellen Gelches, and Jung-Ran Park. “Interoperability and Metadata Quality in Digital Video Repositories: A Study of Dublin Core.” *Journal of Library Metadata* 10, no. 1 (2010): 37–57.