Jon Dieringer

December 14, 2012

Access to Moving Image Collections

CINE-GT 1803

Professor Rebecca Guenther

METADATA CREATION ASSIGNMENT

Arranging Personal, Digital Film-and-Video Collections

PART I: DIGITAL VIDEO COLLECTION

As a cinephile and curator, I maintain a large library of digital video assets. This consists of feature films, short works, video art, advertising clips, educational and industrial films, and other time-based media. Most of them have been transferred from various sources including NTSC and PAL media including VHS, DVD and television broadcast. They are wrapped in various container formats such as AVI, MP4, MOV, MPEG-2, and MKV, and compressed using different video codec libraries and compression standards. The assets have different authors, origins, runtimes, and frame rates among other unique characteristics among both the content they represent and the essence of the media itself. Though I have my own informal system of standardizing file names, it can also be vague or inconsistent—for example, "Ladislas.Starevich.Shorts.1911-1918.avi" or "Entracte 1924 Rene Clair.mp4." Many foreign also have external subtitles such as .srt files.

As this collection grows, I have often thought it would be useful to build a digital asset management system to keep track of these files and attach metadata that would help search the material. For instance, I often obtain screeners of things I may consider programming in the future and store them away unwatched. With a DAM and proper cataloging, if I am curating a show or series particular to a certain time period, artist, genre, or country of origin, I could search into my local digital catalog of materials within these parameters. Or at bare minimum, if

I am looking to spend the evening watching, for instance, Polish avant-garde animation from the 1950s, I could search my holdings much more easily than scrolling through messy, inconsistent file names in my operating system browser.

PART II: METADATA WISH LIST

The following are a list of some qualities that would be useful to record about the digital video collection:

- Record identifier

- Title (e.g. *Citizen Kane*)

- Creator (e.g., John Ford, Mary Ellen Bute, Jam Handy Organization)

- Year

- Runtime (duration of specific digital essence rather than published runtime)

- Country of Origin

- Language

- Description (simple synopsis to pull from a common database like IMDb)

- Various subject tags

- Various genre tags (animation, experimental, video-art, drama, documentary, social-issues, dance)

- File name (Self assigned: "White_Star_1983_Roland_Klick.avi")

- Container format (Mastroka/.mkv, Quicktime/.mov, MPEG-4/.mp4, MPEG-2/.mpeg)

- Codec ID (XviD, ProRes 422, H.264)

- Frame size (1920x1080, 640x480)

- File size (expressed in bytes)

- Encoded frame rate

- Distribution company (whom to contact for screening licenses)

This list is fairly extensive with respect to properties that might concern me personally. However, additional information such as cast and crew members or more esoteric technical information could also potentially be included. There are ways to obtain these automatically; in the case of the former, through an API tied into something like the Open Movie Database, and in the latter, by extracting information through a program like MediaInfo.

III. CATALOGING STANDARDS: DUBLIN CORE, PBCORE and Vocabularies

About Dublin Core

First developed in 1995, Dublin Core contains 15 basic elements which can be adapted across many types of communities. Due to its simplicity and flexibility, it is a good starting place for a personal collection. It would not be intimidating for an amateur cataloger, yet can be expanded as necessary.

Nevertheless, at its most basic, Dublin Core lacks the level of granularity for some of the technical requirements mentioned above. It is difficult to differentiate between things like file size, the essence track duration, container formats and codecs. Therefore, another standard related to these may be preferable. PBCore, below, is based upon Dublin Core.

About PBCore

PBCore is a metadata created by the Corporation for Public Broadcasting containing many more elements and sub-elements. It was first released in 2005, and version 2.0 was published early 2011. It is less rigorous than standards like SMPTE, which contains 1,700 individual elements, and less broadcast-specific than the BBC's SMEF-DM. With respect to those, it is also designed for relative ease of use for the less experienced cataloger beginning with only three required elements: identifier, title, and description. PBCore 2.0 also has several features useful for digital works, such as representing relationships between instantiations (i.e., showing an MPEG4 instantiation encoded from Quicktime instantiation) and the ability to represent multi-part instantiations. It also engages with the semantic web by allowing URI link

attributes for subjects, creators and so on.

PBCore is useful for this collection because it is a smart expansion of Dublin Core tailored specifically to media. It could also be applied to other forms of media in my collection, such as DVDs, while demonstrating potential relationships between them. The ability to specify time-based metadata may also be useful for works that contain various titles, such as one running piece comprised of several short films by an individual artist. Furthermore, the free license for PBCore and rather user-friendly documentation mesh well with this being a personal collection rather than a large institution.

Recommended Vocabularies

The PBCore metadata dictionary suggests vocabularies for many elements, in some cases recommended vocabularies followed by alternates. For instance, the Library of Congress Subject Headings is recommended as an authority for subject headings. This is exhaustive and quite sufficient for the purposes of describing films, and when searching for terms on the Library of Congress website related terms are often suggested. PBCore has its own vocabulary for genres, which is apparently oriented toward broadcast purposes. It lacks the depth or scope to describe many avant-garde, independent, or otherwise non-traditional works—even an over-arching "experimental" term would be helpful. The Getty Thesaurus of Geographic Names is the suggested authority for spatial coverage, and it is of course quite exhaustive and provides useful degrees of granularity. In the case of this collection and its intended use, national names are sufficient. PBCore also recommends ISO 639.2 for 3-letter language codes. For the instantiationDigital element several kinds of video container formats are suggested by PBCore, but I ran into some shortcomings here. The National Archives's PRONOM online registry of technical information was more helpful. Finally, there are some basic and intuitive guidelines for describing things such as frame rates, frame size and duration.

IV. MAPPING COLLECTION TO DUBLIN CORE AND PBCORE

Below, ten "wish list" fields are mapped to DC and PBCore.

FIELD	DUBLIN CORE	PBCORE	NOTES/VOCAB
Title	title	<pbcoreTitle>	PBC: Leading article at end
Director/Creator	creator	<pbcoreCreator> <creator>	May also specify role in PBC PBC recommends AACR2
Year	date	<pbcoreAssetDate>	Both WC3DTF: YYYY
Runtime	format.extent	<pbcoreInstantiation> <instantiationEssenceTrack> <essenceTrackDuration>	PBC: HH:MM:SS
Language	language	<pbcoreInstantiation> <instantiationEssenceTrack> <essenceTrackLanguage>	DC: RFC4646 PBC recommends ISO 639.3 3-letter codes
File Size	format.extent	<pbcoreInstantiation> <instantiationFileSize>	DC: Not very granular PBC: Express in bytes
Container format	format	<pbcoreInstantiation> <instantiationDigital>	PBC recommends MIME but may need to result to other resources. PBC even recommends Wikipedia.
Codec ID	format	<pbcoreInstantiation> <instantiationEssenceTrack> <essenceTrackEncoding>	PBC has extensive vocabulary, eg: ...#h-264-mpeg-4-avc ...#apple-prores-422
Frame Size	format.extent	<pbcoreInstantiation> <instantiationEssenceTrack> <essenceTrackFrameSize>	PBC has its own rules, eg: ...#1280x720
Encoded frame rate	format.extent	<pbcoreInstantiation> <instantiationEssenceTrack> <essenceTrackFrameRate>	PBC has its own rules representing decimals as dashes, eg: ...#23-976 ...#30

Notes about PBCore: Because this collection cataloging exercise is concerned with film titles specifically as digital iterations, aspects such as language and runtime, which may vary from the original release, are specified within the instantiationEssenceTrack element.

V. PBCORE SAMPLE RECORDS

V.i. Feature Film: *DUFFER*

```xml
<PBCoreDescriptionDocument xmlns="http://www.pbcore.org/PBCore/PBCoreNamespace.html"
xmlns:fmp="http://www.filemaker.com/fmpxmlresult" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://www.pbcore.org/PBCore/PBCoreNamespace.html
http://www.pbcore.org/PBCore/PBCoreSchema.xsd">
<pbcoreAssetDate>
        <assetDate>1971</assetDate>
<pbcoreIdentifier>
        <identifier>124123412412341234</identifier>
        <identifierSource>Self-assigned</identifierSource>
</pbcoreIdentifier>
<pbcoreTitle>
        <title>Duffer</title>
</pbcoreTitle>
<pbcoreSubject>
        <subject>
        Homosexuality; Homelessness—Great Britain; Low budget films; Male
        hustlers; Sadomasochism; Voice-overs
        </subject>
        <subjectSource>Library of Congress Headings</subjectSource>
<pbcoreDescription>
        <description>
        Study of a young man who because of his excessive kindness allows
        himself to be badly treated by his male mate.
        </description>
        <descriptionSource>BFI Film & TV Database</descriptionSource>
</pbcoreDescription>
<pbcoreGenre>
        <genre>
```

```
        Adults-Only; Crime; Crime-Drama; Drama; Gay-Lesbian

        </genre>

</pbcoreGenre>

<pbcoreCoverage>

        <coverage>United Kingdom</coverage>

        <coverageSource>TGN</coverageSource>

</pbcoreCoverage>

<pbcoreCreator>

        <creator>Despins, Joseph</creator>

        <creatorRole>Director</creatorRole>

</pbcoreCreator>

<pbcorePublisher>

        <publisher>The Other Cinema</publisher>

        <publisherRole>Distributor</publisherRole>

</pbcorePublisher>

<pbcoreInstantiation>

        <insantiationDigital>AVI (Generic)</instantiationDigital>

        <instantiationDigitalSource>PRONOM</instantiationDigitalSource>

        <instantiationMediaType>Moving-Image</instantiationMediaType>

        <instantiationFileSize>733,545,392</instantiationFileSize>

        <instantiationEssenceTrack>

                <essenceTrackIdentifier>Duffer 1971 Joseph

Despins.avi</essenceTrackIdentifier>

                <essenceTrackEncoding>mpeg-4-xvid</essenceTrackEncoding>

                <essenceTrackFrameRate>25</essenceTrackFrameRate>

                <essenceTrackFrameSize>640x472</essenceTrackFrameSize>

                <essenceTrackDuration>01:12:06</essenceTrackDuration>

                <essenceTrackLanguage>eng</essenceTrackLanguage>

</instantiationEssenceTrack>  </pbcoreInstantiation>  </PBCoreDescriptionDocument>
```

V.ii. Short Film*: Debt Begins at 20*

```
<PBCoreDescriptionDocument xmlns="http://www.pbcore.org/PBCore/PBCoreNamespace.html"

xmlns:fmp="http://www.filemaker.com/fmpxmlresult" xmlns:xsi="http://www.w3.org/2001/XMLSchema-

instance" xsi:schemaLocation="http://www.pbcore.org/PBCore/PBCoreNamespace.html

http://www.pbcore.org/PBCore/PBCoreSchema.xsd">

<pbcoreAssetDate>

        <assetDate>1980</assetDate>

<pbcoreIdentifier>

        <identifier>124123412412341235</identifier>

        <identifierSource>Random</identifierSource>

</pbcoreIdentifier>

<pbcoreTitle>

        <title>Debt Begins at Twenty</title>

</pbcoreTitle>

<pbcoreSubject>

        <subject>

        Documentary-style films; Experimental films; Feminism and music;

        Feminist motion pictures; Fictionalized documentary films; Personal

        films; Pseudo-documentary; Punk culture in motion pictures; Punk rock

        musicians; Punk rock music; Quasi-documentary films; Semidocumentary

        films; Underground films

        </subject>

        <subjectSource>Library of Congress Headings</subjectSource>

<pbcoreDescription>

        <description>

        "DEBT BEGINS AT TWENTY, by combining semi-fictionalized and documentary

        material, is as definitive a record of the Pittsburgh punk scene during

        its nascent underground phase as anyone could possibly hope for. That

        it also succeeds as an extremely engaging character study of some of
```

```
        the movement's most colorful figures is a testament to Stephanie

        Beroes' facility for combining a series of 'miniatures' with empathy as

        well as to the unusually natural character of the participants in what

        elsewhere has been an excruciatingly arch subculture." - W.T. Koltek,

        WYEP-Radio, Pittsburgh

        </description>

        <descriptionSource>Canyon Cinema Catalog</descriptionSource>

</pbcoreDescription>

<pbcoreGenre>

        <genre>

        Comedy-Drama; Concert; Docudrama; Documentary; Entertainment; Gay-

        Lesbian; Music

        </genre>

</pbcoreGenre>

<pbcoreCoverage>

        <coverage>United States</coverage>

        <coverageSource>TGN</coverageSource>

</pbcoreCoverage>

<pbcoreCreator>

        <creator>Beroes, Stephanie</creator>

        <creatorRole>Director</creatorRole>

</pbcoreCreator>

<pbcorePublisher>

        <publisher>Canyon Cinema</publisher>

        <publisherRole>Distributor</publisherRole>

</pbcorePublisher>

<pbcoreInstantiation>

        <insantiationDigital>FLV</instantiationDigital>

        <instantiationDigitalSource>Library of Congress Sustainability of
```

```
Digital Formats</instantiationDigitalSource>

        <instantiationMediaType>Moving-Image</instantiationMediaType>

        <instantiationFileSize>199,062,539</instantiationFileSize>

        <instantiationEssenceTrack>

                <essenceTrackIdentifier>Debt Begins at 20 1980 Stephanie

Beroes.avi</essenceTrackIdentifier>

                <essenceTrackEncoding>vp6</essenceTrackEncoding>

                <essenceTrackFrameRate>25</essenceTrackFrameRate>

                <essenceTrackFrameSize>480x360</essenceTrackFrameSize>

                <essenceTrackDuration>00:36:27</essenceTrackDuration>

                <essenceTrackLanguage>eng</essenceTrackLanguage>

        </instantiationEssenceTrack>

</pbcoreInstantiation>

</PBCoreDescriptionDocument>
```

V.iii. Experimental Short: *Lapis*

```
<PBCoreDescriptionDocument xmlns="http://www.pbcore.org/PBCore/PBCoreNamespace.html"

xmlns:fmp="http://www.filemaker.com/fmpxmlresult" xmlns:xsi="http://www.w3.org/2001/XMLSchema-

instance" xsi:schemaLocation="http://www.pbcore.org/PBCore/PBCoreNamespace.html

http://www.pbcore.org/PBCore/PBCoreSchema.xsd">

<pbcoreAssetDate>

        <assetDate>1966</assetDate>

<pbcoreIdentifier>

        <identifier>124123412412341236</identifier>

        <identifierSource>Random</identifierSource>

</pbcoreIdentifier>

<pbcoreTitle>

        <title>Lapis</title>

</pbcoreTitle>

<pbcoreSubject>

        <subject>

        Abstract Films; Animation; Computer Animation; Avant-Garde Films;

        Computer-Assisted Filmmaking; Computer-Generated Animation;

        Experimental Films; Nonobjective Films; Nonrepresentational Films

        </subject>

        <subjectSource>Library of Congress Headings</subjectSource>

<pbcoreDescription>

        <description>

        Mathematically precise early computer-aided animated piece based around

        the meditative-like inward and outward flow of images of the mandala.

        </description>

        <descriptionSource>Internet Movie Database</descriptionSource>

</pbcoreDescription>

<pbcoreGenre>
```

```
    <genre>

    Animation; Computers; Music

    </genre>

</pbcoreGenre>

<pbcoreCoverage>

    <coverage>United States</coverage>

    <coverageSource>TGN</coverageSource>

</pbcoreCoverage>

<pbcoreCreator>

    <creator>Whitney, James</creator>

    <creatorRole>Director</creatorRole>

</pbcoreCreator>

<pbcoreInstantiation>

    <essenceTrackIdentifier>Lapis 1966 James

Whitney.avi</essenceTrackIdentifier>

    <insantiationDigital>AVI (Generic)</instantiationDigital>

    <instantiationDigitalSource>PRONOM</instantiationDigitalSource>

    <instantiationMediaType>Moving-Image</instantiationMediaType>

    <instantiationFileSize>136,300,544</instantiationFileSize>

    <instantiationEssenceTrack>

        <essenceTrackEncoding>mpeg-4-xvid</essenceTrackEncoding>

        <essenceTrackFrameRate>25</essenceTrackFrameRate>

        <essenceTrackFrameSize>720x576</essenceTrackFrameSize>

        <essenceTrackDuration>00:09:17</essenceTrackDuration>

        <essenceTrackLanguage>No Dialog</essenceTrackLanguage>

    </instantiationEssenceTrack>

</pbcoreInstantiation>

</PBCoreDescriptionDocument>
```