Metadata in Long-Term Digital Preservation

Grzegorz Płoszajski

Abstract Metadata has been used to describe digital data for several years. First metadata standards were created in the nineties. Particularly important proved to be Dublin Core designed for the description of web-based resources (1995). At the same time Exif standard was created, determining mainly technical metadata used to describe parameters of digital images and information on equipment used. Since that time various other types of metadata have been created, e.g. rights metadata, provenance, structural, and many domain-oriented standards. Long-term digital archives usually preserve all metadata transferred to them with corresponding digital objects and often use a subset of obtained metadata to manage archive assets. They also create preservation metadata, which describe processes and preservation actions applied to digital objects in the archives. In this chapter basic information about metadata in long-term digital archives is given.

1 Introduction

1.1 Digital Preservation

The goal of digital preservation is to keep digital materials accessible and usable for long periods of time. Digital information is fragile. It can be lost because of media degradation. Another problem is technological obsolescence caused by introducing new software and new technologies, which supersede older. The older fall out of use.

Digital preservation must be performed actively. Preservation activities can be applied to stored digital information in institutional repositories, digital libraries and archives; hereinafter referred to collectively as digital archives. Archives apply such strategies as data redundancy, migration and emulation. Data redundancy (making several copies of digital objects) may be helpful in short-term. Migration can be done to new formats of files or to new media; it solves to some extent problem

G. Płoszajski (✉)
Institute of Control and Computation Engineering, Warsaw University of Technology, Warsaw, Poland
e-mail: G.Ploszajski@ia.pw.edu.pl

T. Traczyk et al. (eds.), *Digital Preservation: Putting It to Work*,
Studies in Computational Intelligence 700, DOI 10.1007/978-3-319-51801-5_2

of technological obsolescence. Migration is addressed to long-term preservation. Emulation can help when new software supersedes older; files do not have to be migrated to new formats. But emulation software must be updated.

1.2 Preservation and Metadata

Generally it is a good idea to think about preserving digital objects when they are created. One of possible activities is gathering as much information about each digital object as possible from the start. Such information is called metadata.

Common practice in digitization of cultural heritage is creating metadata right after creating corresponding digital objects. In such case metadata should contain information about original object and about the created digital one (e.g. parameters of digital image, equipment used, data and time of creation etc.). Metadata can contain supplementary information concerning quality check of created object (e.g. description of flaw on an image) or process of additional transformations of object (e.g. elimination of skew). It should contain legal information concerning access to digital object and allowed, and/or forbidden transformations. Metadata is called according to nature of information, e.g. descriptive, technical, administrative, rights, provenance.

Many digital objects are created for current utility purposes; the need to preserve these objects may become conscious after some time. Personal collections of photo images might be an example. Creating metadata in such case is more difficult and for some types of metadata even impossible.

When digital objects together with corresponding metadata are 'sent' to digital archive to be preserved for a long time, process of preservation starts. Another type of metadata is used to keep record of preservation activities made in the archive. This group is called preservation metadata. An archive attempting to preserve digital data for long time should use preservation metadata. When digital object is returned to the owner after a long time it should be accompanied by the metadata, which will inform the user of its content.

Systems of metadata are usually defined in a formal way. They are often called standards (some even passed long way to become ISO standards). Most of metadata systems use XML notation. XML Schema can be treated like a simple standard definition. Some systems of metadata are described with use of dictionaries. One of first standards of preservation metadata is PREMIS, managed by the Library of Congress.

1.3 Short and Long-Term Preservation

Digital Preservation Handbook Glossary [1] developed by Digital Preservation Coalition defines short, medium and long-term preservation as follows:

- *Short-term preservation*—Access to digital materials either for a defined period of time while use is predicted but which does not extend beyond the foreseeable future and/or until it becomes inaccessible because of changes in technology.
- *Medium-term preservation*—Continued access to digital materials beyond changes in technology for a defined period of time but not indefinitely.
- *Long-term preservation*—Continued access to digital materials, or at least to the information contained in them, indefinitely.

In practice terms "short-term" and "medium-term" are met rather seldom, however the case where information data should be preserved only until some defined period of time can happen often, usually when there is a legal obligation to store some kind of documents (e.g. master thesis in Poland should be preserved during 50 years).

There is also another group of information which should be stored to guarantee usability of digital data (content and metadata) after long time but is not treated as metadata. This can be standards of metadata and dictionaries of terms used in metadata. Without them after a long time information contained in metadata might be difficult to understand.

2 Metadata in Digital Archives

2.1 Digital Archive and Customers

For the purpose of this chapter we consider model of digital archiving containing two elements:

- digital archive *and*
- customers.

Customers entrust their digital data to the archive and archive takes responsibility for keeping these data usable for long time, and provide access to the data for the customers. This may be university repository and scientific staff of a university or business oriented digital archive offering services and external customers, individual or companies.

The data can be transported to the archives either by network or on physical media, e.g. hard disks or CDs. In the second case the archive reads and store the data from the media. In both cases there should be a procedure according to which customer can make sure that digital data in the archive is an exact copy of the original.

In such model it is usually the customer's responsibility to provide metadata; the data sent to the archive should consist of the 'content' data (e.g. digital images, sound, text, measurement data) and metadata. Archive becomes responsible only for keeping metadata sent by customers and creation of preservation metadata.

Responsibility of the archive and the customer concerning the 'content' data and metadata can be formally based on agreements between them. Agreements may state accepted formats of files and standards of metadata as well as procedures of delivery

of data to the archive (SIP) and of sending it back to the user after a "long" time (DIP).

Metadata is often external to the content file—saved in separate file—however can also be embedded in the content files, especially technical metadata.

The model consisting of archive and customers does not fit to web archiving, where the content of the web is often grabbed as it is. Creators of the web content are neither customers nor partners of the archive. They do not have to adhere to metadata standards. They can use different standards. They do not have to create metadata at all. The archive must manage such inconsistent information. We shall not consider such model in this chapter.

2.2 Metadata Received and Metadata Created in the Archive

In the context of processing data in long-term archives metadata can be divided into two groups:

- metadata supplied to the archive together with related digital objects,
- metadata created in the archive.

There are various types of metadata supplied to archives, e.g. descriptive metadata, technical, structural, administrative, rights, provenance, business [2]. Technical metadata can contain parameters of digital image or audio file. Structural metadata can show how a complex object is built and what is a role of individual data files. Structural metadata can also represent structure of SIP packages sent to the archive.

Some of the above mentioned metadata are created manually, especially description metadata, some are created by hardware, e.g. technical metadata of images in photo cameras, and some by software, e.g. measurement metadata of scientific experiments or space missions.

Second group contains preservation metadata, which make record of operations made in the archive, e.g. transformation of SIP to AIP, counting checksums, comparing checksums of different copies of individual files, migration to new file formats or migration to new memory media. These metadata should be created by software used for management of archived data.

However some metadata from first group can also be included in preservation metadata. Digital objects can be created as 'born-digital' or as digital 'copies' of physical (analogue) objects, e.g. images of paper documents or museum artifacts. Such objects can be processed, e.g. to eliminate skew or noise or simply to change format (e.g. RAW to TIFF, Word to PDF/A). History of such operations should be recorded in provenance metadata. Provenance information may be important in the future for assessing authenticity of digital objects and to make judgments about whether information is trusted. (Term *provenance* is discussed in detail in [3].)

Generally, provenance metadata should be gathered by the customers from the time of creating the object and transferred to the archive; provenance information can be included in preservation metadata during ingest and supplemented during preservation activities.

Metadata sent to archive can be stored separately from the content files, however in long-term archives metadata is usually treated as part of digital object; all metadata can be subject to preservation activities together with related content files.

When digital object comes to an archive, part of its metadata is often copied to a database (or other data structure) to help in management of archive's assets. Part of metadata transferred to management database should contain:

- all or some of descriptive metadata to help in searching objects,
- a small number of technical metadata to help make decisions concerning migration to new formats,
- rights metadata to control access to stored objects or to information on these objects.

2.3 Evolution of Metadata Standards

2.3.1 New Versions of Standards

Metadata standards evolve. They get new versions, e.g. standard VRA Core [4], used for the description of images and work of art and culture, created in the late nineties, got version 4.0 in 2007; ISO 19115 standard DIF [5] used in earth sciences, got version 6.0 in 2010; METS [6], created in 2001, got version 11.0 of XML Schema in 2015. As a result metadata based on different versions of metadata standards must co-exist in long-term digital archives and archives must be able to manage assets containing metadata based on the variety of standard versions.

A number of data dictionaries and ontologies have been developed as a part of metadata standards to help in preparing consistent metadata (e.g. PREMIS [7]). New standard versions get new dictionaries and ontologies.

Archives should store metadata standards with these dictionaries and ontologies (as well as taxonomies and thesauri). After many years they should help to understand what is the real meaning of some metadata. They might also help in precise searching.

2.3.2 Development of Dublin Core

Dublin Core is domain-agnostic standard, created during series of workshops which began in 1995 in Dublin, Ohio. Simple Dublin Core Metadata Element Set(DCMES) has 15 elements: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights. Each element is optional and repeatable. Descriptive names 'promote' a common semantic understanding of the elements. List was published in 1998 [8]. Using controlled vocabularies for specific elements could increase interoperability of the standard.

A number of domain-oriented metadata standards was inspired by this list of elements, e.g. EBU Core metadata set for Radio archives, Tech 3293, version 1.0, 2001 [9] (EBU—European Broadcasting Union [10]). Some standards got mappings

from Dublin Core, e.g. Directory Interchange Format DIF used in Earth Science [11], or to Dublin Core.

Another achievement was creation by *Open Archives Initiative* Protocol for Metadata Harvesting (OAI-PMH) [12], based on obligatory use of Simple Dublin Core.

Dublin Core has been used with types of materials which demanded some complexity, e.g. linking dates with events or persons with functions. As result of discussions qualifiers have been developed. Since then the two versions of Dublin Core were distinguished: the older one called *simple* or *unqualified* and the new one called *qualified* with qualifiers [13].

Qualifiers could be used to refine an element. For example element *Date* got five refinements: *Created, Valid, Available, Issued, Modified* and element *Relation* twelve refinements: *Is Version Of, Has Version, Is Replaced By, Replaces, Is Required By, Requires, Is Part Of, Has Part, Is Referenced By, References, Is Format Of, Has Format*. Qualifiers could also be used to indicate encoding scheme. For example element *Subject* got five controlled vocabularies (i.a. LCSH, MeSH, UDC), element *Language*—two standards: ISO 639-2 abd RFC 1766.

Qualifiers did not solve all problems. Standard was changed. New version is called DCMI Metadata Terms [14]. 55 elements have been defined: (*abstract, accessRights, accrualMethod, accrualPeriodicity, accrualPolicy, alternative, audience, available, bibliographicCitation, conformsTo, contributor, coverage, created, creator, date, dateAccepted, dateCopyrighted, dateSubmitted, description, educationLevel, extent, format, hasFormat, hasPart, hasVersion, identifier, instructionalMethod, isFormatOf, isPartOf, isReferencedBy, isReplacedBy, isRequiredBy, issued, isVersionOf, language, license, mediator, medium, modified, provenance, publisher, references, relation, replaces, requires, rights, rightsHolder, source, spatial, subject, tableOfContents, temporal, title, type, valid*). In the dictionary every element has URI identifier in corresponding namespace using persistent uniform resource locators (http://purl.org/dc/terms/).

Dublin Core was published as ISO Standard 15836 in 2009. Not all problems of users, which inspired changes, have been resolved by this version. Ability to link persons to their roles was improved, nevertheless was not satisfactory for many users. Standard DCMI Metadata Terms allows to link agents to roles from closed list, being a small subset (about 30 elements from 150) of *MARC Code List for Relators* [15].

Dublin Core was defined without using XML notation. However there are published guidelines for implementing Dublin Core in XML [16]), and validating with XML Schemas: for Simple Dublin Core [17] and for DC Terms [18].

There is a number of digital libraries which use Dublin Core as descriptive metadata. dLibra software, which is popular in Poland, allows to use simultaneously Simple Dublin Core and other descriptive metadata standards (using Simple Dublin Core is obligatory). OAI-PMH Protocol (*Open Archives Initiative Protocol for Metadata Harvesting* [12] allows to harvest metadata from libraries or archives, which also use this protocol. Some large archives make use of protocol OAI-PMH, i.a. arXiv [19]. In Poland metadata from all dLibra libraries and some others using OAI-PMH, are harvested to the central library of Digital Libraries Federation FBC [20] and made available for searching.

The above mentioned OAI-PMH protocol has created own namespace and own XML Schema, based on DCMI schema for unqualified Dublin Core.

2.3.3 Development of Domain-Oriented Metadata Standards

Domain oriented metadata standards have been subject to similar development as Dublin Core and changed a lot. However motivation to making changes usually was different.

One example is EBU Core, which in 2001 in first version with its 15 elements was almost identical to Simple Dublin Core [9]. In 2015 version 1.6 of standard was constructed in totally different way [21]. Changes introduced in version 1.6 are explained in introduction to specification in the following way: "EBUCore 1.6 takes into account latest developments in the Semantic Web and Linked Open Data communities. EBUCore 1.6 is available as a RDF ontology entirely compatible with the W3C Media Annotation Working Group ontology, which model is common and based on the EBU Class Conceptual Data Model (Tech.3351). A link to the RDF/OWL ontology and its documentation is provided in Annex B. The EBUCore ontology has been updated to complement EBU's CCDM (Tech 3351) and improve mapping with other ontologies. EBUCore RDF is listed as Linked Open Vocabulary as well as RDF-Vocab for Ruby developers. The definitions in EBUCore 1.6 have been refined. The schema structure has been reinforced for registration in EBU's Class 13 in SMPTE. The new advanced data model for audio defined in Tech 3364 and introduced in EBUCore 1.5 has been updated to reflect discussions around its adoption in ITU." Thus creation of version 1.6 was motivated not only by 'internal' EBU problems but also to enhance business cooperation of European producers with US (SMPTE—Society of Motion Picture and Television Engineers [22] is an American organization setting standards for Motion Imaging) and to be in touch with developments taking place in International Telecommunication Union [23] (ITU is the United Nations specialized agency for information and communication technologies ICTs).

Interesting history have Photo metadata standards created by International Press Telecommunications Council (IPTC) [24] to improve exchange of news among newspapers and agencies (see Metadata History Timeline and Guidelines). Metadata can be embedded in photo files. IPTC Photo Metadata standard is supported by such software as Adobe Photoshop.

Great role of Simple Dublin Core remains also valid in version DCMI Metadata Terms.

2.4 Minimal Set of Metadata

Archives can define requirements concerning data and metadata delivered to them. Sometimes this is done in form of defining minimal requirements.

American National Archives and Record Administration (NARA) applied such approach to metadata. NARA published in 2015 *Metadata Guidance for the Transfer of Permanent Electronic Records* [25].

Recommendation is addressed to federal agencies. They must prepare metadata which accompany transfers of permanent electronic records to national archives. Besides descriptive, technical and administrative information concerning structure and content of electronic records "metadata elements also provide contextual information that explains how electronic records were created, used, managed, and maintained prior to their transfer to NARA, and how they are related to other records. This information enables NARA to appropriately manage, preserve, and provide access to electronic records for as long as they are needed."

NARA defined minimum metadata requirements as a subset of the Dublin Core Metadata Element Set v.1.1.

"Agencies should provide the following elements for each file or item included in a transfer (…):

1. Identifier [File Name]. The complete name of the computer file including its extension (if present);
2. Identifier [Record ID]. The unique identifier assigned by an agency or a records management system;
3. Title. The name given to the record;
4. Description. A narrative description of the content of the record, including abstracts for document-like objects or content descriptions for audio or video records;
5. Creator. The agent primarily responsible for the creation of the record;
6. Creation Date. The date that the file met the definition of a Federal record; and
7. Rights. Information about any rights or restrictions held in and over the record including access rights such as national security classification, or personally identifiable information, Privacy Act, or Freedom of Information Act, or usage rights relating to copyright or trademark.

Agencies should provide the following metadata elements, if they apply to the record being transferred:

1. Coverage. The geographic and temporal extent or scope of the content of the record; and
2. Relation. The relation element should be used if a record is composed of multiple files that form a logical record, or is a necessary component of another logical record.

If an agency provides additional metadata elements, NARA will accept that metadata as part of the transfer process in addition to NARA minimum metadata requirements. Agencies should notify NARA of any metadata standards that are in use with permanent electronic records and provide relevant schemas, data dictionaries, controlled vocabularies, ontologies, and system indexes at the time of transfer."

For the element Rights five refinements were given (four new; only RightHolder was defined in DCMI Metadata Terms):

- Security Classification—mandatory,
- Previous Security Classification—mandatory if applicable,
- Access Rights—mandatory,
- Usage Rights—mandatory when applicable,
- Rights Holder—mandatory when applicable.

2.5 Comments on File Formats

Some repositories are oriented to achieve two goals: provide short-term access and long-term usability. To imagine what "short time" can be, consider problems one might have with reading text documents prepared with word processing programs 25 years ago. Certain file formats might become not accessible for current applications. To read texts from this archive an extra software may be needed. In case of short-term archives it is rather the users' problem to find and use such software, while long-term archives should convert files, which might become obsolete to new formats. Such operation is called format migration in digital archives.

Migration is rather costly. Archives try to limit number of file formats which they accept.

Analysis of digital formats with respect to their usability in long-term digital preservation have been made for long time by many organizations. The Library of Congress manages the gathered results [26] using title *Sustainability of Digital Formats Planning for Library of Congress Collections*. There is also given comprehensive information concerning evaluation factors taken into account, such as:

- Sustainability factors—apply for all categories of information:
  - Disclosure,
  - Adoption,
  - Transparency,
  - Self-documentation,
  - External Dependencies,
  - Impact of Patents,
  - Technical Protection Mechanisms;
- Quality and functionality factors (for selected content types):
  - Still Images:
    Clarity (support for high image resolution),
    Color maintenance (support for color management),
    Support for graphic effects and typography,
    Support for multispectral bands;

- Sound:
  Fidelity (support for high audio resolution),
  Support for multiple channels (including note-based, e.g., MIDI),
  Support for downloadable or user-defined sounds, samples, and patches;
- Text:
  Support for integrity of document structure and navigation,
  Support for integrity of layout, font, and other design features,
  Support for rendering for mathematics, formulas, diagrams, etc.;
- Moving Images:
  Clarity (support for high image resolution),
  Fidelity (support for high audio resolution),
  Support for multiple sound channels.

Some archives make lists of preferred file formats or of accepted file formats. Good example of such practice is shown in a *Revised Format Guidance for the Transfer of Permanent Electronic Records* [27]. American National Archives and Record Administration (NARA) defined in this document ten categories and a few subcategories of digital resources:

1. Computer Aided Design,
2. Digital Audio,
3. Digital Moving Images:
   a. Digital Cinema,
   b. Digital Video,
4. Digital Still Images:
   a. Digital Photographs,
   b. Scanned Text,
   c. Digital Posters,
5. Geospatial Formats,
6. Presentation Formats,
7. Textual Data (plain text, formatted text, word processed),
8. Structured Data Formats (databases, spreadsheets, scientific data),
9. Email,
10. Web Records.

Seven categories are common with the Library of Congress list [26] (category *Datasets* can be treated as equivalent to *Structured Data Formats*). Three: *Computer Aided Design, Presentation Formats* and *Email* have no equivalent categories. The Library of Congress defines also eighth category *Generic*, designed for wrappers (e.g. RIFF), bundling formats (which can be used to construct SIPs and DIPs) and encodings.

For each of ten categories NARA declared formats using up to three levels of format acceptance:

- Preferred Formats,
- Acceptable Formats,
- Acceptable for Imminent Transfer Formats.

The third level is designed for "legacy formats that are no longer in common use and that NARA will eventually stop accepting". Special procedure of transfer of such files is required. In most cases only two initial categories of formats are defined. The third one is given only for two categories: 5 and 8.

Decisions made by American archives can be inspiring for other archives.

Migration of file formats can have implications for metadata. If metadata is external, the information on change of content file format can be stored in preservation metadata and in Data Management; there is no need to change original metadata. In case of embedded metadata, migration of content to new format may lead to problems if new format is unable to embed original metadata. However, every archive which accepts files with embedded metadata should be prepared to preserve metadata. At least it can export embedded metadata to external file at the time of ingest.

3 Metadata of Typical Digital Archive Resources

3.1 Categories of Digital Resources

Digital archives gather data coming from various domains of life. Description of those data is usually domain-oriented and there are numerous standards of descriptive metadata. Categories of digital objects are less diverse, because they can be used in many domains. For example still images can be used to make digital copies of analogue originals such as paper documents and museum artifacts, and to create and save "born digital" objects such as medical documentation or satellite photographs.

Problems concerning long-term preservation of digital images are similar regardless of domain of the application. In Sect. 2.5 are shown ten categories of digital objects, formulated by the American National Archives [27]. Categories listed by the Library of Congress in [26] are shown below:

- Still Image,
- Sound,
- Textual
- Moving Images,
- Web Archive,
- Datasets,
- Geospatial,
- Generic.

The first seven from this list is common with the list made by National Archives; the first four are discussed further in this section.

3.2 Domain-Oriented Metadata Standards

A number of domain-oriented metadata standards have been created. List of about 40 metadata standards was published by Digital Curation Centre [28] and list of about 35 metadata standards in Wikipedia [29]. Some universities have published similar lists, e.g. list of Domain Metadata Standards published by University of Central Florida Libraries [30]. None of those lists aspire to be complete.

Below there is given a list of metadata standards to show the variety of domains for which metadata standards have been created:

- AgMES—Agricultural Metadata Element Set (FAO); used also as namespace [31].
- AVM—Astronomy Visualization Metadata (International Virtual Observatory Alliance), standard used to tag astronomical images [32].
- PROV—W3C specification [33] designed for interchange of provenance information.
- VRA Core—a data standard for the description of images and works of art and culture (designed by Visual Resources Association 1996 [4], version 4.0 hosted by the Library of Congress [34]).
- MODS—descriptive metadata standard designed for libraries, which allows for more detailed description of resources than Dublin Core [35].
- EBUCore—EBU Core Metadata Set, which describes audio and video resources for a wide range of broadcasting applications [21].
- DIF—Directory Interchange Format—for exchanging information about scientific data sets [5].
- TEI: Text Encoding Initiative—a standard for the representation of texts in digital form, mainly in the humanities, social sciences and linguistics [36].
- ONIX—a family of metadata standards: for Books, Serials and Licensing Terms & Rights Information, designed originally to support book trade, extended to "support creating, distributing, licensing and otherwise making available intellectual property in published form, whether physical or digital" [37].
- STEP—Standard for the Exchange of Product Model Data, describing how to represent and exchange digital product information, and manage the information during life cycle of the product (officially "Industrial automation systems and integration—Product data representation and exchange" a group of standards ISO 10303) [38].

Metadata standards can be designed with a view to support long-term preservation of information. An interesting discussion concerning "Engineering-specific Metadata Requirements for Long-term Archiving" in the context of Digital Product Data Archiving and metadata standard STEP are presented in [39].

3.3 Mappings Between Metadata Standards

There is a need to make mappings between metadata standards. Several such mappings have been defined. Library of Congress and W3C have published some of them.

- MARC—Dublin Core: documents concerning the so called "crosswalk" of library standard MARC21 to Dublin Core (2001) [40] and both versions of Dublin Core—DCMES and DCMI Metadata Terms—to MARC (2008) [41].
- MODS—Dublin Core: mapping version 3 was defined in both directions on standard MODS Official Website in 2012:
  - MODS to Dublin Core—there is a comment: "In some cases, multiple MODS elements are equivalent to one Dublin Core element, and all are listed. In these cases, some metadata is lost."
  - Dublin Core to MODS—there is a comment: "In some cases, the Dublin Core element is equivalent to more than one MODS element, and all are listed. In these cases, when converting a record from Dublin Core to MODS, a decision may need to be made as to the default MODS element."
- Dublin Core (DCMES) to CIDOC CRM [42]—Conceptual Reference Model CIDOC [43] is a domain ontology in the area of cultural heritage, created by the International Council of Museums (ICOM), developed into an ISO standard 21127.
- Dublin Core to PROV mapping published by W3C [33]. PROV is a provenance metadata standard published by W3C. This mapping is made between the PROV-O OWL2 ontology and the Dublin Core Terms Vocabulary [14].

Interesting problem of mapping metadata concerns cooperation of cultural heritage institutions from various countries with Europeana [44]. Europeana uses specific format—the Europeana Data Model EDM. Europeana offers five possible routes of supplying data [45]; three of them are based on mapping. Each institution can use existing crosswalk to map and transform data directly to EDM, or to an intermediary schema or to one of standard metadata schemas, such as Dublin Core or EAD, and map to EDM using special tool (MINT). However, Europeana is oriented to sharing data and giving access on mass scale and not to long-term archiving.

3.4 Still Images

Digital still images is a basic category of assets in many digital archives. They can be used in all cases where the classic photography can be used. Digital images may play a role of copies of various paper documents, stored usually in libraries and archives, or copies of photographic images captured on films, glass plates or photographic paper. Similar status of digital copies of analogue originals may have photographies

of museum artifacts. In the above mentioned cases term digitization is usually used; digital images are made by scanners and photographic cameras.

In the archives there are also stored digital photographies made for scientific, medical, commercial and other purposes. Some of those images are created by means of photographic cameras, some are results of computer analysis and transformation of measurement data.

Generally these images can be treated in similar way to ensure long-term preservation. Some differences, e.g. time of migration to new formats, may result from different formats of stored digital images.

3.4.1 Dublin Core and Other Descriptive Metadata Standards

Images can be described in Dublin Core metadata. Many digital libraries uses this standard for images. In Poland many digital libraries which are using dLibra software, describe images in DC.

Dublin Core is not well suited to describe images so some archives, libraries and museums use other metadata standards or own systems. Polish National Digital Archive (Narodowe Archiwum Cyfrowe) offers search form where user besides elements from DC have such fields as: DateFrom, DateTo, Persons visible and Persons not visible. Museums often describe their holdings in much more detailed way than DC makes possible. Some metadata can be even collection-dependent.

Librarians in previous years used to describe all their collections in MARC, even if it was difficult, e.g. in case of music records. Now they can use MODS instead [35]. Due to possibility of mapping metadata standards, e.g. MARC to MODS and MODS to DC, such institutions can offer users descriptions in two or three standards, among them is the Library of Congress.

As an example will be used descriptions of a map *New railroad map of the state of Maryland, Delaware, and the District of Columbia* available at Library of Congress [46] available in MODS and Dublin Core; MODS record is about three times as big as Dublin Core record.

Listing 1 Description of a map in Dublin Core standard (Library of Congress)

```
<dc:title>
  New railroad map of the state of Maryland,
  Delaware, and the District of Columbia. Compiled
  and drawn by Frank Arnold Gray.
</dc:title>
<dc:creator>Gray, Frank Arnold.</dc:creator>
<dc:subject>
  Railroads--Middle Atlantic States--Maps.
</dc:subject>
<dc:description>
  Shows drainage, canals, stations,
  cities and towns, counties, canals, roads completed,
  narrow gauge and proposed railroads with names
  of lines. Includes list of railroads.
</dc:description>
<dc:description>Scale 1:633,600.</dc:description>
<dc:description>LC Railroad maps, 230</dc:description>
<dc:description>
        Description derived from published bibliography.
```

```
</dc:description>
<dc:publisher>Philadelphia</dc:publisher>
<dc:date>1876</dc:date>
<dc:type>image</dc:type>
<dc:type>map</dc:type>
<dc:type>cartographic</dc:type>
<dc:identifier>
  http://hdl.loc.gov/loc.gmd/g3791p.rr002300
</dc:identifier>
<dc:language>eng</dc:language>
<dc:coverage>
  United States--Middle Atlantic States
</dc:coverage>
```

Listing 2 Fragments of description of the same map in MODS which have no equivalent information in DC

```
...
<originInfo>
  <place>
   <placeTerm type="code" authority="marccountry">pau</placeTerm>
  </place>
  ...
</originInfo>
<language>
  <languageTerm type="code" authority="iso639-2b">eng</languageTerm>
</language>
<physicalDescription>
  <form authority="marccategory">electronic resource</form>
  <form authority="marcsmd">remote</form>
  <extent>col. map 39 x 62 cm.</extent>
</physicalDescription>
...
<note type="additional physical form">
  Available also through the Library of Congress Web site
  as a raster image.
</note>
...
<recordInfo>
  <recordContentSource authority="marcorg">DLC</recordContentSource>
  <recordCreationDate encoding="marc">980520</recordCreationDate>
  <recordChangeDate encoding="iso8601">20021214203459.0</recordChangeDate>
  <recordIdentifier>5572072</recordIdentifier>
  <recordOrigin>
    Converted from MARCXML to MODS version 3.5 using
    MARC21slim2MODS3-5.xsl (Revision 1.106 2014/12/19)
  </recordOrigin>
</recordInfo>
```

Last part of additional information contained in MODS can be treated as provenance metadata.

3.4.2 Standard IPTC Photodata Core and Extension

IPTC published version 1.2 of IPTC Photo Metadata standard in 2014 [47]. However, version 1.1 of IPTC Core and IPTC Extension from 2010 is still widely used [48].

Standard was designed to be used by photographers, news agencies and publishers. It is supported by camera software of many producers and by professional photo editing software. The standard is also widely used by the amateurs.

Structure of metadata standard is simple list of terms without hierarchy and repetition. IPTC Core has four sections: *Contact, Content, Image* and *Status*:

- IPTC Contact:
  - Creator (name of the person who created the photograph),
  - Creator's Job Title (e.g. staff photographer),
  - Contact Info (Address…Phone, Email, Website);
- IPTC Image:
  - Date Created (date the photograph was created),
  - Intellectual Genre (term from *IPTC Genre Newscode*),
  - IPTC Scene Code (term from *IPTC Scene-Newscodes*),
  - Location (Sublocation, City, State/Province, Country, ISO Country Code);
- IPTC Content:
  - Headline (short synopsis of the content of photograph),
  - Description (who, what, where, when…),
  - Keywords,
  - IPTC Subject Code (IPTC Subject Newscodes taxonomy),
  - Description Writer;
- IPTC Status:
  - Title (short reference—text or numeric),
  - Job ID (can be added for transmission purposes),
  - Instructions,
  - Credit Line (how the owner should be credited when the image is published),
  - Source (original owner or copyright holder),
  - Copyright Notice (current owner or copyright holder),
  - Rights Usage Terms (how this photograph can be legally used).

IPTC Extensions defines five groups of terms: *Description, Artwork/Object, Models, Administrative* and *Rights*, which can make the description more detailed. For example the *Location* has now two fields and two meanings: *Location Created* in which the image was created and *Location Shown* in the image. IPTC Core and Extension examples are shown on page of IPTC [49]. However description of image can be limited to IPTC Core only.

In Poland this standard is used by some news agencies and commercial photographers; also by such public institutions as Museum of History of Photography [50] and History Meeting House [51].

Listing 3 IPTC metadata of exemplary photography of Museum of History of Photography (translated)

```
Caption             :
Date Created        : 1915-05-1915-08
Keywords            : MAN/PRIVATE LIFE - everyday life
  - uniform, WAR/ARMY - warfare, TYPES OF PRESENTATION
  - Portrait - group portrait
Object Name         : Two legionnaires, group portrait
```

```
Copyright Notice    : public domain, www.mhf.krakow.pl
Byline              : unknown author
```

Listing 4 IPTC metadata of exemplary photography of History Meeting House (translated)

```
Keywords            : Brandel Konrad, Gembarzewski Leszek
   - collection,  tenements, church, Warsaw,
Caption             : DI 36164; Brandel, Konrad
   (1838-1920) (photographer);  Warsaw; about 1895;
        photography; photographic paper; 11,8 x 16,3
Program             : FotoWare FotoStation
Byline              : Ligier Piotr
Caption Writer      : MM-A
Copyright Notice    : Copyright by Ligier Piotr/National Museum Warsaw
```

The IPTC Core metadata can be embedded in image files such as JPEG, TIFF and PSD either in older IIM (Information Interchange Model) data format or in newer Adobe's XMP data format. The newer one can be also embedded in DNG and PDF files.

3.4.3 Exif Standard—Technical Metadata

Exif Standard—Exchangeable Image File Format was created by the Japan Electronic Industries Development Association in 1995 (v. 1) and modified several times. Versions 2.3 and its 2.31 revision were published in 2010 and 2016 by two associations: Japan Electronics and Information Technology Industries Association and Camera & Imaging Products Association, nevertheless versions 2.2 (2002) and 2.21 (2003) are still being used in many products. A detailed information concerning revision history is enclosed in standard specification: *CIPA DC-008-Translation-2016. Exchangeable image file format for digital still cameras: Exif Version 2.31* [52].

Exif standard provided method of recording data in image files at the point of image capture; the method was defined for two file formats: JPEG (ISO/IEC 10918-1) and TIFF v. 6.0. The standard has been used in digital photo cameras and scanners, but had no provisions for information specific for scanners. Standard covers a broad spectrum of data: camera settings (such as shutter speed, focal length), date and time, thumbnail, copyright and description.

Total number of Exif metadata increased in version 2.2 to 147 [53]; in this number there were also 31 GPS tags. If camera had GPS receiver (internal or external) the GPS tags could be captured by camera software. They could also be added later by external software.

Below there are shown Exif metadata read from a photography made by SLR Sony DSLR-A700 in internal format RAW and transformed by camera software Image Data Converter SR to TIFF format and to JPEG format.

Listing 5 Exif metadata of photography made by SLR as RAW and saved in TIFF format

```
[Camera]
Camera Manufacturer         : SONY
Camera Model                : DSLR-A700
Software                    : Image Data Converter SR
Date modified               : 2015:09:18 09:06:00
```

```
[Image]
Exposure time [s]          : 1/500
F-Number                   : 5.6
Exposure program           : Aperture priority (3)
ISO speed ratings          : 200
EXIF version               : 02.21
Date taken                 : 2014:07:22 14:15:26
Date digitized             : 2014:07:22 14:15:26
Brightness                 : 203/25
Exposure bias value        : 0
Max aperture               : F2.8
Metering mode              : Spot (3)
Light source               : Daylight (1)
Flash                      : No flash
Focal length [mm]          : 24
User comment               :
FlashPix Version           : 01.00
Colour space               : sRGB
Custom rendered            : Normal process (0)
Exposure mode              : Auto (0)
White balance              : Manual (1)
Digital zoom               : 0
Focal length (35mm)        : 36
Scene capture type         : Standard (0)
Contrast                   : Normal (0)
Saturation                 : Normal (0)
Sharpness                  : Normal (0)
```

Exif metadata of the same photography, saved in JPEG format, contains more information.

Listing 6 Additional Exif metadata of the same photography saved by the camera in JPEG format

```
Components configuration   : YCbCr
Compressed bits per pixel  : 8
EXIF image width           : 4272
EXIF image length          : 2848
Interoperability offset    : 25460
File source                : DSC
Scene type                 : A directly photographed image
```

Standard Exif was extended in version 2.2 to cover the sound registered by photo camera. The Exif audio file specification defines method of writing audio data in files. The method was defined for RIFF WAVE Form Audio File format only. As data format is used Pulse Code Modulation (PCM) for uncompressed audio data (and also G.711, used mainly in US and Japan) and Adaptive Differential Pulse Code Modulation (IMA-ADPCM) for compressed audio data.

3.4.4 Standard MIX

Exif metadata are usually presented in form of lists similar to shown above. Originally standard MIX did not use XML notation. There have been made some efforts, e.g. Exif vocabulary workspace—RDF Schema [54] in 2003, but there was no official XML Schema.

Library of Congress published *MIX—NISO Metadata for Images in XML Schema. Technical Metadata for Digital Still Images Standard* [55]; NISO stands here for National Information Standards Organization [56]. Standard is addressed to raster

digital images. MIX Schema version 2.0 was published in 2008. Data Dictionary—Technical Metadata for Digital Still Images (ANSI/NISO Z39.87-2006) is available as NISO Standard.

The MIX standard contains Exif metadata from versions 2.2 and 2.21. It has also small number of preservation metadata concerning fixity of stored image files (group *ChangeHistory* with subgroups of metadata *ImageProcessing* and *PreviousImageMetadata*. It can be applied to such file formats as JPEG2000, DjVu and MrSID for which Exif can't be applied.

MIX standard can be used to store Exif metadata extracted from image files. It might be useful in case of migration to new formats or could be done at the time of ingest just for safety of embedded metadata (the last motivation is due to the fact that popular software sometimes corrupts Exif metadata even at such simple operations as copying). XML notation is not necessary for storing extracted metadata, however formal structure controlled by XML schema might ensure correctness of such operation.

3.4.5 Exif and IPTC Embedded in XMP

Format TIFF 6.0 [57] has provided space for Exif metadata and IPTC metadata (IIM). It can also contain Extensible Metadata Platform (XMP) [58] metadata.

It should be noted that there is trend in photo camera industry to use XMP on greater scale. XMP can be used as a container for traditional Exif and IPTC metadata. It can also be used to contain precise information concerning camera settings technical parameters of photographs—much more precise than it was possible when using Exif metadata.

To illustrate ability of recording metadata in XMP below are cited Exif and IPTC metadata, and small fragments of metadata in XMP of the same photography: CopyrightInFotos-MDTest01a.jpg, published by IPTC [49].

Listing 7 Exif metadata of exemplary photography

```
[Camera]
Date modified              : 2012:04:22 20:07:21
Y Resolution               : 240
Software                   : Adobe Photoshop Lightroom 4.0 (Windows)
Image description          : Bikefestival in Wien, Rathausplatz
Camera Manufacturer        : Canon
Camera Model               : Canon EOS 60D
Copyright                  : Copyright 2012 Frank Fotofan
                             www.ffotofan.info
Artist                     : Frank Fotofan
Resolution unit            : Inch
Orientation                : top-left (1)

[Image]
Exposure bias value        : 0
EXIF version               : 02.30
Shutter speed [s]          : 1/250
Focal length [mm]          : 100
Date digitized             : 2011:04:03 13:16:34
```

```
Subject distance (m)       : 25
SubSecTimeDigitized        : 34
F-Number                   : 5.6
Focal plane Y-Resolution   : 518400/119
Focal plane X-Resolution   : 777600/181
White balance              : Auto (0)
Max aperture               : F4.2
Aperture                   : F5.6
Focal plane res. unit      : Inch (2)
Metering mode              : Multi-segment (5)
Flash                      : No flash
Exposure program           : Normal (2)
Custom rendered            : Normal process (0)
Scene capture type         : Standard (0)
```

Listing 8 IPTC metadata of exemplary photography

```
Date Created               : 20110403
Headline                   : Bikefestival
Caption                    : Bikefestival in Wien,
                             Rathausplatz
Copyright Notice           : Copyright 2012 Frank Fotofan
                             www.ffotofan.info
Credits                    : Frank Fotofan
Location                   : Rathausplatz
City                       : Wien
Country                    : Oesterreich / Austria
```

Listing 9 Fragment 1 of XMP metadata of exemplary photograph in RAW format

```
<rdf:Description rdf:about=""
      xmlns:aux="http://ns.adobe.com/exif/1.0/aux/">
  <aux:SerialNumber>0380227035</aux:SerialNumber>
  <aux:LensInfo>70/1 300/1 0/0 0/0</aux:LensInfo>
  <aux:Lens>EF70-300mm f/4-5.6L IS USM</aux:Lens>
  <aux:LensID>489</aux:LensID>
  <aux:LensSerialNumber>000000617b</aux:LensSerialNumber>
  <aux:ImageNumber>0</aux:ImageNumber>
  <aux:ApproximateFocusDistance>251/10</aux:ApproximateFocusDistance>
  <aux:FlashCompensation>0/1</aux:FlashCompensation>
  <aux:Firmware>1.0.5</aux:Firmware>
</rdf:Description>
```

Listing 10 Fragment 2 of rights XMP metadata of exemplary photograph in RAW format

```
<rdf:Description rdf:about=""
      xmlns:xmpRights="http://ns.adobe.com/xap/1.0/rights/">
  <xmpRights:Marked>True</xmpRights:Marked>
  <xmpRights:WebStatement>
    http://creativecommons.org/licenses/by-nc-sa/3.0/de/
  </xmpRights:WebStatement>
  <xmpRights:UsageTerms>
    <rdf:Alt>
      <rdf:li xml:lang="x-default">Creative Commons - by-nc-sa/3.0/at/</rdf:li>
    </rdf:Alt>
  </xmpRights:UsageTerms>
</rdf:Description>
```

Listing 11 Fragment 3 of XMP metadata of exemplary photograph in RAW format

```
<rdf:Description
      rdf:about=""xmlns:crs="http://ns.adobe.com/camera-raw-settings/1.0/">
  <crs:Version>7.0</crs:Version>
  <crs:ProcessVersion>5.7</crs:ProcessVersion>
```

```
  <crs:WhiteBalance>As Shot</crs:WhiteBalance>
  <crs:Temperature>5000</crs:Temperature>
  <crs:Tint>+1</crs:Tint>
  <crs:Exposure>0.00</crs:Exposure>
  <crs:Shadows>5</crs:Shadows>
  <crs:Brightness>+50</crs:Brightness>
  <crs:Contrast>+25</crs:Contrast>
  <crs:Saturation>0</crs:Saturation>
  <crs:Sharpness>25</crs:Sharpness>
  <crs:LuminanceSmoothing>0</crs:LuminanceSmoothing>
  <crs:ColorNoiseReduction>25</crs:ColorNoiseReduction>
  ...
  <crs:ToneCurve>
    <rdf:Seq>
       <rdf:li>0, 0</rdf:li>
       <rdf:li>32, 22</rdf:li>
       <rdf:li>64, 56</rdf:li>
       <rdf:li>128, 128</rdf:li>
       <rdf:li>192, 196</rdf:li>
       <rdf:li>255, 255</rdf:li>
    </rdf:Seq>
  </crs:ToneCurve>
  <crs:ToneCurveRed>
    ...
  </crs:ToneCurveRed>
  ...
</rdf:Description>
```

The examples presented above show great ability of XMP (using also RDF) to record with great precision parameters of photographs and camera settings, beyond the scope of Exif standard.

However, this ability does not improve anything with respect to long-term archiving. Format RAW—as commercial—is not accepted in this role, the more so there is no one RAW format but many different formats of various producers.

Digital Negative (DNG) format can be used in the archives. It is an open lossless format, written by Adobe [59], designed for archiving RAW photographs taken by cameras of various manufacturers. Photographs saved in different versions of RAW format can be converted to one DNG format. Using this DNG format and converters is royalty free. DNG format contains checksum information important in archiving. In the Format Guidance for the Transfer of Permanent Electronic Records [27] the DNG format (specification version 1.4.0.0) has status of acceptable format for category *Digital Photography*.

3.5 Sound Archives

3.5.1 Introduction

Sound archive or audio archive can be a part of collections of a library, an archive or museum, e.g. British Library Sound Archive. Sound archive can also be a separate institution, e.g. radio archive. Sound archives from many countries created in 1969 *International Association of Sound Archives (IASA)* [60], which in 2012

extended range of collections, changing name to *International Association of Sound and Audiovisual Archives* but leaving the the abbreviation IASA unchanged.

IASA published i.a. the *IASA Cataloguing Rules (1999)* [61] and the *Guidelines on the Production and Preservation of Digital Audio Objects* (2009) [62].

3.5.2 IASA Cataloging Rules

The IASA Cataloging Rules define 11 groups of metadata elements:

0: Preliminary notes,
1: Title and statement of responsibility,
2: Edition, issue, etc.,
3: Publication, production, distribution, broadcast, etc., and date(s) of creation,
4: Copyright,
5: Physical description,
6: Series,
7: Notes,
8: Numbers and terms of availability,
9: Analytic and multilevel,
10: Item/copy information.

Physical description concerns physical media, such as tapes, records and disks; large part of archives' resources were using such media in 1999.

Each of 11 groups contains several elements, e.g. group no 3 *Publication, production, distribution, broadcast, etc., and date(s) of creation* contains:

3.0. Scope and definitions,
3.A. Preliminary rule,
3.B. General rule,
3.C. Place of publication, production, distribution, broadcast, etc.,
3.D. Name of publisher, producer (production company), distributor, broadcaster, etc.,
3.E. Optional addition. Statement of function of publisher, producer (production company), distributor, broadcaster, etc.,
3.F. Date of publication, production, distribution, broadcast, etc.,
3.G. Place, name and date of manufacture,
3.H. Date(s) of creation (unpublished items only),
3.I. Reproductions (which are themselves unpublished).

Each of above elements has brief description and explanation.

From 37 examples given by IASA, with various types of content, like popular music, classical music, opera, oral history, spoken word, radio music etc. on various types of media, the following two have been chosen:

Example 1. Radio music production on DAT—live recording

Johannes Brahms und seine Freunde (1)/Markus Brändle, producer; Erich Heigold, sound engineer; Gertrud Bastuck, cut.—Saarbrücken: SR2, 26.12.1988, (13.05–4.30).—Copyright: SR; GEMA; GVL.—1 sound tape (DAT, 84 min 46 s): digital (AAD), stereo, AES/EBU standard.—(Der musikalische Salon) Rolf Sudbrack, author and speaker; Jewgenij Koroliov, Ljupka Hadzigeorgiev, piano. Recorded live 27.01.1988, Saarbrücken (Germany), Funkhaus Halberg (Grosser Sendesaal) Sampling frequency and quantisation: 48 kHz, 16 bit. Contents: Variationen über ein Thema von Robert Schumann, Es-dur. op.23 (Leise und innig) / Johannes Brahms (1 min 49 s)—4 Balladen op.10 (Nr.3 h-moll; Nr.4 H-dur)/Johannes Brahms (4 min 30 s.; 9 min 36 s.)—3 Romanzen op.21 (Nr.1 a-moll; Nr.2 F-dur)/Clara Schumann (5 min 28 s; 1 min)—4 Klavierstücke op.2 (Nr.2 Kanon)/Albert Dietrich (4 min 20 s)—Sonate für Klavier zu 4 Händen g-moll, op.17 (2.Satz)/Hermann Goetz (5 min 33 s)—Intermezzo B-dur, op.76 Nr.4/Johannes Brahms (2 min 29 s)—Capriccio d-moll, op.116 Nr.7/Johannes Brahms (2 min 10 s). Copy from 2 tape reels, analogue: stereo; 38 cm/s, Telcom C4.

Example 2. Opera film

Aida [videorecording] / music: Giuseppe Verdi; original libretto: Antonio Ghislanzoni; produced by: Staffan Rydn; directed by: Claes Fellbom; revised text and screenplay: Claes Fellbom; costume designer: Inger Pehrsson; art director: Lotta Melanton; choreography: Ann-Charlotte Lindström.—Sweden: the Swedish Film Institute [distributor]: Isis Film, the Swedish Film Institute, Sveriges television TV2 and Filmhuset [production companies], [1993].—1 videocassette (VHS, ca. 122 min): sd. (stereo), col. Credits: Berndt Fritiof/Filmmixarna, sound/mix supervisor; Hans Ewers, sound/music supervisor; Jörgen Persson, director of photography; et al. Cast: Margareta Ridderstedt (Aida), Niklas Ek (Radames), Robert Grundin (Radames' voice), Ingrid Tobiasson (Amneris), Jan van der Schaaf (Amonasro), Alf Häggstam (Ramfis), Staffan Rydén (Pharao's spokesman), et al. The Swedish Folkopera Orchestra and Choir; Kerstin Nerbe, conductor. Filmed version of Verdis Aida, shot in Lanzarote, Spain. This production was originally staged by the Swedish Folkopera.

In the above examples descriptive metadata are used together with technical metadata and sometimes also provenance metadata. Relations between persons or institutions and their roles, e.g.

Jewgenij Koroliov, Ljupka Hadzigeorgiev—piano;
Giuseppe Verdi—music;
the Swedish Film Institute—distributor;
Jörgen Persson—director of photography (credits);
Niklas Ek—Radames (cast);

represent significant part of information concerning the archives' holdings.

3.5.3 Acceptable Formats for Audio

The Format Guidance for the Transfer of Permanent Electronic Records [27] contains following recommendations concerning preferred or acceptable formats for Digital Audio:

- Preferred formats:
  - Broadcast Wave (BWF); codec LPCM (Linear Pulse Code Modulated Audio); version 1 and version 2; EBU Tech specification 3285,
  - Free Lossless Audio Codec (FLAC); version 1.21.

- Acceptable formats:
  - Audio Interchange Format (AIFF); codec LPCM (Linear Pulse Code Modulated Audio); v. 1.3,
  - MPEG Audio Layer III (MP3); codec MP3enc; ISO/IEC-11172-3 part 3—Audio [63],
  - Waveform Audio File Format (Wave); codec LPCM (Linear Pulse Code Modulated Audio).

General technical recommendations: sample rate at least 44.1 kHz, but 96 kHz is encouraged; minimum of 16 bits per sample, but 24 bits per sample is encouraged.

3.5.4 Audio Metadata Standards

New edition of *Guidelines on the Production and Preservation of Digital Audio Objects* (2009) [62] is based on the Reference Model for an Open Archival Information System (OAIS). Discussion and recommendations concerning metadata take into account aspects of preservation, e.g. "the record of the creation of the digital audio file, and any changes to its content, must be created at the time the event occurs. This history metadata tracks the integrity of the audio item and, if using the BWF format, can be recorded as part of the file as coding history in the BEXT chunk. This information is a vital part of the PREMIS preservation metadata recommendations." In these guidelines there are mentioned such metadata standards as Simple Dublin Core, DCMI Metadata Terms with Applications Profiles, AES57 (Audio Engineering Society), METS and PREMIS. Full name of AES57 is: *AES57-2011: AES standard for audio metadata—Audio object structures for preservation and restoration*. Standard provides vocabulary.

Relations between persons and their roles can hardly be expressed in Dublin Core. It is suggested in the Guidelines to use MARC Code List for Relators [15] and DCMI Terms, as in the example below:

```
<dcterms:contributor>
  <marcrel:CMP>Giuseppe Verdi</marcrel:CMP>
  <marcrel:CNG>J\"orgen Persson</marcrel:CNG>
</dcterms:contributor>
```

Verdi is tagged here as composer (CMP), Persson as cinematographer (CNG), what is equivalent to director of photography. However, not all roles mentioned in the above examples are on the relators list, e.g. there is no piano (pianist etc.), only general term musician. Expressing casting in Dublin Core is even more difficult. IASA Guidelines suggest that in such cases MODS is a better standard than DC.

In this context EBU Core Metadata Standard [21] can also be taken into account. In the first version this standard [9] (2001) EBU decided to add refinements to elements from simple Dublin Core list. One of the refinements was *Role* which was added to elements: *Creator* and *Contributor*. The EBU Reference Data Tables: Roles in broadcasting was also added to these elements as recommended Encoding scheme for *Role*, with the following *EBU Comment*: "The element refinement Role is added

by EBU and not part of standard DCMES. The content of the qualifier Role must be taken from a controlled list of authorized roles. It is recommended that roles are taken from the EBU Reference Data Table [6], but this list can be extended to cover special local needs."

EBU Comments were added to all elements. Element *Subject* got a comment: "Persons as subjects are also placed here. Genre of the content is placed under element Type"; a few comments concern "Recommended best practice". A few *Element Encoding schemes* were added, among them *EBU Reference Data Table: Type of resource* and *RDS: PTY display terms* for element *Type* and *SMPTE Unique Material Identifier (UMID), International Standard Recording Code (ISRC)* and *International Standard Audiovisual Number (ISAN)*.

New version of EBU Core [21] (2015) recommends using EBU Classification Schemes and ontologies [64]. Roles are defined in *EBU Role Code Classification Scheme* (2013) [65]. List contains about 680 terms—much more than list of *MARC Code List for Relators*.

3.5.5 Formats and Embedded Metadata

Basic recommended format for archiving the sound is WAVE (extension WAV). It is designed to store audio bitstream in chunks (a specific container format). Most often WAV files contain uncompressed sound in the Linear Pulse Code Modulation (LPCM) format, however are also able to contain compressed sound. WAV can contain metadata (tags) in the INFO chunk; it can embed any kind of metadata.

There is also version of WAV called BWF (Broadcast Wave Format), specified by EBU in 1997 as EBU Tech 3285, updated in 2001 (when SMPTE UMID was added) and revised in 2011. BWF has extension chunk BEXT designed for broadcasters needs. Format is specified for use in AES31 and also by SMPTE 382M (Mapping of AES and Broadcast Wave audio into the MXF generic container). In 2012 BWF became able to contain identifiers, especially the International Standard Recording Code ISRC (EBU Tech 3352: The Carriage of Identifiers in the Broadcast Wave Format BWF).

The MP3 specification is part of the MPEG-1 standard (1992) designed for lossy compression of video and audio. MPEG-1 standard is published as ISO/IEC 11172—*Information technology—Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s*. Part III of the standard, which concerns Audio, specifies three layers: layer I,layer II and Layer III. MP3 is an abbreviation of *MPEG-1 Layer III*. MP3 is a lossy compression of audio and as such is not recommended to use for long-term archiving. However it became extremely popular and it may happen that some MP3 files will be sent to long-term archives (if there is no uncompressed version). MP3 specification does not define tag formats, however there are *de facto* metadata standards. Metadata are usually embedded at the end or at the beginning of MP3 file.

In version 1 of the standard last 128 bytes of file contained 6 fields, which might be converted to Dublin Core:

- Song Title—30 bytes—Title,

- Artist—30 bytes—Contributor,
- Album—30 bytes—Source,
- Year—4 bytes—Date,
- Comment—30 bytes—Description,
- Genre—1 byte—Subject (to be decoded).

In 1998 was created Version 2 (ID3v2) of the standard, which introduced radically different structure of metadata. Metadata information is added at the start of the file, not at the end. Fields (tags) are of variable size, not fixed. Each field has own identifier (four chars in versions 2.3 and 2.4). Many fields were added; total number increased to 74 in version 2.3 and 83 in version 2.4. In most of these fields there is descriptive and technical metadata. There are also fields for commercial information and for user defined text information.

Information from basic descriptive tags can be easily mapped to Dublin Core, e.g.

- MCDI—Music CD Identifier—to Identifier;
- TALB—Album/Movie/Show title—to Source;
- TCOM—Composer—to Contributor (Role);
- TCOP—Copyright message—to Rights;
- TDRC—Recording time—to Format;
- TEXT—Lyricist/Text writer—to Contributor (Role);
- TIT2—Actual name of the piece (e.g. “Adagio”)—to Title;
- TLAN—Language—to Language;
- TPE1—Lead artist/Lead performer/Soloist—to Contributor;
- TPE3—Conductor—to Contributor (Role);
- TPUB—Publisher—to Publisher;
- TSRC—ISRC (International Standard Recording Code)—to Identifier;
- WCOP—Copyright/Legal information—to Rights.

Standard ID3 was designed to be used in MP3, but can also be used in other formats, i.a. WAV and MP4.

3.6 Cinema and TV Archives

3.6.1 Introduction

FIAF—Fédération Internationale des Archives du Film/International Federation of Film Archives [66] was founded in 1938 in Paris by four film archives from US, UK, France and Germany. In 2015 FIAF had 155 affiliates (85 members and 70 associates) in 74 countries. TV archives were created later, mainly by TV broadcasters. European Broadcasting Union had no special forum for TV archive issues. A few archives which recognized their need of collaboration and sharing expertise created in 1976 IFTA—International Federation of Television Archives [67].

The need of various forms of cooperation among audiovisual archives motivated creation of the *Co-ordinating Council of Audiovisual Archives Associations* (CCAAA) which was founded at the end of the 1990s. Besides FIAF and IFTA the participants of CCAAA were the International Council of Archives (ICA), the International Association of Sound Archives (IASA), and the International Federation of Library Associations (IFLA).

Among fields of common interest were cataloging and (later) metadata standards, which were earlier developed within libraries and 'traditional' archives.

3.6.2 FIAF Cataloging Rules

FIAF published Cataloging Rules for Film Archives in 1991 [68]. These rules were based on libraries experience in bibliographic description: International Standard Bibliographic Description for Non-Book Materials (ISBD (NBM)). The approach to describing films predates the era of digital production. It is focused on 'reels' of film and their localization in the archive. Below there is exemplary description shown in [68].

Listing 12 Exemplary description of film based on 1991 rules (fragments)

```
TITLE          The Princess' Necklace
STATEMENTS OF RESPONSIBILITY
               director, Floyd France ;
               story, Clare Freeman Alger ;
               scenarist, E. Clement D'Art
PRODUCTION     US : Thomas A. Edison, Inc.
DISTRIBUTION [producer], 1917 ;
ETC.           US : K.E.S.E. [distributor], 1917
COPYRIGHT      (c): US: Thomas A. Edison, Inc. 31 Aug 17;
               LP11335.
               Viewing print: 4 reels of 4(1498 ft.):
               16 mm.: S., b&w, si. / USW FLA 1742-1745.
               Duplicate negative: 4 reels of 4(1498 ft.);
PHYSICAL       16mm. : S., b&w, si. / USW FRA 4336-4339.
DESCRIPTION Archival positive: 4 reels of 4(1498 ft.);
               16mm. : S., b&w, si. / USW FRA 4340-4343.
SERIES         (CONQUEST PROGRAM ; NO. 8)
               Cast: William Calhoun, Kathleen Townsend,
               Wallace MacDonald, Susan Mitchell,
               Dorothy Graham, Roy Adams.
NOTES          Summary: A fairy tale in which a stranger
               comes to Happyland in order to learn the
               master secret of happiness (...)
               References: Moving Picture World, (...)
```

New approach to cataloging films, based on metadata, was published by FIAF scarcely in 2016 in *The FIAF Moving Image Cataloging Manual* [69]. Film archives couldn't wait from 1991 to 2016. They had to implement other standards. Authors of this document made comments concerning relationship of new document to standards which were applied in the meantime. One comment refers to European Standards Committee standards EN 15744 and 15907. s comment concerns i.a. the EBU standard: "While these guidelines are intended to be applicable to all forms of moving image materials, archives with extensive broadcasting collections may wish to

look to broadcasting-specific metadata schemas such as EBUCore and PBCore for additional guidance."

3.6.3 CEN European Standards Committee

European Standards Committee (CEN) published two Cinematographic Works Standards: EN 15744 and EN 15907 which define the metadata essential for facilitating data exchange and consistent identification of moving images [70]. Standards are available at [71, 72].

European metadata standard EN 15744 defines following minimum set of metadata for cinema films:

- Title,
- Series/Serial,
- Cast,
- Credits,
- Production Company,
- Country of Reference,
- Original Format,
- Original Length,
- Original Duration,
- Original Language,
- Year of Reference,
- Identifier,
- Genre,
- Relationship,
- Source.

Simplicity of this list reminds that of Dublin Core, but there are differences. The meaning of each data element is explained in [73]. Three elements relate to physical parameters of "the first known manifestation of a cinematographic work" (original format, length, duration).

Metadata standard 15907 [74] has more complex data model which allows to make much more detailed descriptions.

- Primary Entities (Cinematographic Work, Variant, Manifestation, Item, Content).
- Contextual Entities (Agent, Event).
- Elements (Identifier, Record Source, Title, Identifying Title, Country of Reference, Year of Reference, Format, Extent, Language, Production Event, Publication Event, Award, Decision Event, IPR Registration, Preservation Event, Subject Terms, Content Description).
- Common Element Types (Region, Timespan, Language Tag).
- Relationships (HasAgent, HasEvent, HasContent, HasAsSubject, HasOtherRelation, HasVariant, HasManifestation, HasItem).

3.6.4 Comment on FIAT/IFTA Cataloging Rules

FIAT/IFTA published in 1992 a minimum list of 22 fields for cataloging broadcast materials [75]. The list contained three groups of fields: descriptive fields (i.a. identification, personnel information, keywords, time and place of shooting), technical (physical description, date of transmission) and legal. List was available for members of FIAT/IFTA only.

3.6.5 Metadata in TV Archives—EBU Survey

Standard EBUCore is used by a number of TV archives. The EBUCore was discussed in Sect. 3.5.4 as standard concerning audio assets. However this standard concerns both audio (radio) and TV assets. It can be easily noticed e.g. in *EBU Role Code Classification Scheme* (2013) [65], where list of codes was created by joining the MPEG7 cast roles plus TV-Anytime Role Classification System [76] (TV-Anytime is also managed by EBU).

EBUCore was also indicated in the new FIAF Moving Image Cataloguing Manual [69]. It could be done because FIAF extended scope of Manual from films to TV. Together with EBUCore was indicated standard PBCore—Public Broadcasting Metadata Dictionary, developed in US to describe sound and moving images, digital and analog. Authors claim that: "Because it is so useful in describing media assets, a growing number of film archives and media organizations outside of public broadcasting have adopted PBCore to manage their audiovisual assets and collections [77]." Like EBUCore, the PBCore is based on Dublin Core. Version 2.1 of PBCore was released in 2015.

Another standard should be mentioned here: P/META standard [78]—managed by EBU— "originally designed to support business to business content exchanges, it has also been implemented for other purposes like for exchange between production systems or as high level descriptive semantic metadata".

A number of metadata standards for TV is registered in SPTME Metadata Registry class 13–14 [79] managed by SMPTE Registration Authority, LLC. Class 13 is reserved for metadata registered by an organization for public use. Among 14 organizations which registered metadata in this class are: EBU, Pro-MPEG Forum, Association of Moving Image Archivists, Public Broadcasting Service (PBS), Audio Engineering Society (AES) and Library of Congress.

EBU prepared an *Archive Report 2010* [80] based on a survey concerning TV archive-specific issues. One of them concerned the metadata standard.

Archives were divided into two groups:

- archives with integrated file-based facilities—advanced,
- archives with partially file-based facilities—beginners.

Two main groups of users were Broadcasters and Vendors. Two groups of metadata standards were standards for exchange and standards for internal use. Main results of the survey:

- Metadata standards for exchange:
  - In the group of archives with *integrated file-based facilities* "almost 20% of Broadcasters used Dublin Core based standards for exchange between archive and production" and 13% used their own internal metadata scheme. The majority of Vendors used schemes according to user requirements.
  - "Broadcasters with *partially file-based facilities* either specify their own in-house formats (most popular) or use a Dublin Core based or P/META based Metadata scheme for exchange."
- Metadata standards for internal use:
  - "Broadcasters with *integrated file-based facilities* indicate mainly the use of their own proprietary formats (most popular—25%) or use Dublin Core based formats (13%)"; small number of archives use P/META based Metadata. The majority of Vendors used schemes according to user requirements.
  - "Broadcasters with *partially file-based facilities* report similar results" as for exchange.

Authors of the survey conclude: "In-house formats are more frequently specified for internal archive usage, whereas Dublin Core based formats serve mainly for Metadata exchange between archive and production."

3.6.6 Comments on File Formats

According to [27] there is given one preferred format for Digital Cinema, i.e. Digital Moving Picture Exchange Bitmap (DPX), specified by SMPTE 268M; accepted formats are not specified in this case.

As to Digital Video, preferred formats are not specified in [27]; there are indicated following accepted formats:

- Audio Video Interleaved Format (AVI)—uncompressed 4:2:2,
- QuickTime File Format (MOV)—uncompressed 4:2:2,
- Windows Media Video 9 File Format (WMV)—codec: VC-1,
- MPEG 4—codec: H.264,
- MPEG-2 Video (MPEG2),
- Material Exchange Format (MXF)—J2K-losslessly-compressed.

In Poland Apple ProRes 422 HQ is also recommended; MXF is not recommended (not popular).

4 Rights Metadata

Many metadata standards are able to contain information concerning intellectual rights to digital objects. For example Dublin Core has a field *Rights* in which simple information without qualifiers can be stored. More complex rights information can be

stored in IPTC, but this standard is rather unsuitable to be used outside the commercial photography.

Rights metadata are needed in many domains. Various standards of rights metadata have been created. Among worth mentioning is Open Digital Rights Language (ODRL) [81], being developed under auspices of W3C (in 2015 version 2.1 of Core Model, XML Encoding, Common Vocabulary and Ontology). As an example of commercial standards can be indicated group of standards created by consortium DDEX [82]. Consortium defined group of over 20 standards DDEX used to support exchange of information among the media companies, music licensing organizations, digital service providers etc. However, the long-term preservation does not need such detailed information concerning rights.

Among standards deserving more attention in context of long-term archiving are:

- METS, which can be used to create packages SIP, and is designed to contain rights metadata;
- PREMIS, which defines characteristics of rights concerned with preservation activities and gathers information during storage of digital objects in the archive.

Rights metadata in standard METS are defined in external XML Schema: METSRights.xsd [83]. There are three main elements defined:

- RightsDeclaration—description of intellectual property rights associated with digital object or its part,
- RightsHolder—a person or organization which owns some intellectual property rights,
- Context—a description concerning who has what permissions and constraints.

In the file METSRights.xsd are also given comments concerning metadata and examples. Below is one of examples:

Listing 13 Exemplary metadata in standard METSRights (fragments)

```
<RightsDeclaration>
  Any re-use of these materials in publication may
  only be done with the explicit permission of the
  Charles L. Dodgson Estate. Please contact the
  Fales Library staff if you wish to use any of
  these materials.
</RightsDeclaration>
...
<RightsHolder RIGHTSHOLDERID="FALESRH01">
  <RightsHolderName>The Estate of Charles L.
    Dodgson (Lewis = Carroll)</RightsHolderName>
  <RightsHolderComments>
   The estate of Charles Dodgson is represented by
   AP Watt agency of London. All permissions issues must
   be addressed to them.
  </RightsHolderComments>
  <RightsHolderContact>
    ...
  </RightsHolderContact>
</RightsHolder>
...
<Context CONTEXTCLASS="GENERAL PUBLIC">
  <Permissions OTHER="false" PRINT="false"
    DELETE="false" MODIFY="false"
```

```
    DUPLICATE="false" COPY="false"
    DISPLAY="true" DISCOVER="true"/>
  <Constraints CONSTRAINTTYPE="QUALITY">
    <ConstraintDescription>
      Users may only access digital copies of photographic
      materials digitized at 50 dpi or less.
    </ConstraintDescription>
  </Constraints>
</Context> <Context CONTEXTCLASS="REPOSITORY MGR">
  <Permissions OTHER="false" PRINT="true"
    DELETE="true" MODIFY="true"
    DUPLICATE="true" COPY="true"
    DISPLAY="true" DISCOVER="true"/>
</Context>
```

The above informations concern the current situation. There is some lack of information concerning problems which might appear in the future with respect to preservation activities.

Standard PREMIS is based on entity semantic units: Object, Event, Agent and Rights.

Data Dictionary version 2.2 describes the *Rights Entity* using following terms:

Listing 14 Rights metadata for Entity Semantic Units in standard PREMIS (Data Dictionary version 2.2)—chosen terms

```
rightsStatement
rightsStatementIdentifier
rightsBasis
copyrightStatus
copyrightJurisdiction
copyrightDocumentationIdentifier
copyrightApplicableDates
startDate
endDate
licenseDocumentationIdentifier
licenseDocumentationRole
licenseTerms
licenseApplicableDates
statuteInformation
statuteJurisdiction
statuteCitation
statuteDocumentationIdentifier
statuteApplicableDates
otherRightsInformation
rightsGranted
act
restriction
termOfGrant
termOfRestriction
linkingObjectIdentifier
linkingObjectRole
linkingAgentIdentifier
linkingAgentRole
rightsExtension
```

This short list gives an idea what could be expressed by such terms.

Choice between PREMIS and METS in case of rights metadata should be made carefully. Library of Congress published Guidelines for using PREMIS with METS [84], which may be helpful.

5 Delivery of Content Data and Metadata to Archive

5.1 Packages and Metadata Standards

The OAIS reference model defines three types of information packages:

- Submission Information Package (SIP)—information delivered from the content provider to the archive,
- Archival Information Package (AIP)—information stored by the archive,
- Dissemination Information Package (DIP)—information delivered to a user on request.

Generally packages contain:

- digital objects, i.e. content files and metadata files (or embedded metadata) concerning objects or individual files, and
- metadata concerning the package itself.

Metadata concerning the package should inform at least about the structure of the package: e.g. list of files, formats of files, structure of catalogs etc. and about such attributes of files as checksums. Some such system of metadata became standards. It usually means that they have been defined in detailed and precise way and have syntax controlled by XML Schema which is maintained by an organization responsible for the standard.

Several such 'packaging' metadata standards have been developed, among them:

- METS: Metadata Encoding and Transmission Standard, designed for libraries (version 1.1 in 2002) [6],
- XFDU: XML Formatted Data Unit, designed for space data (ISO standard 13527 in 2003) [85],
- LOTAR: Metadata for Archival Package Workgroup [86],
- E-ARK: European Archival Records and Knowledge Preservation [87].

METS and XFDU achieved greater significance than other two and are presented in this chapter in more details.

METS was used in many projects. METS Implementation Registry [88] contains list of about 40 projects from 8 countries. List of METS registered profiles [89] extends this list further.

Information concerning the use of XFDU is gathered in different way [90]. CCSDS gathers information about missions which are "known to be using CCSDS-recommended protocols". Showing list of over 800 such missions CCSDS claims that "many of these missions also follow CCSDS Recommendations for data archiving" what can mean that they use XFDU. Among projects which are known for using XFDU two European projects can be mentioned:

- project CASPAR [91]—Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval—an EU Integrated Project,

- project SAFE—Standard Archive Format for Europe [92].

Requirements concerning packages are not limited to metadata standard. They become more specific for individual projects. The project KOPAL can be used as an example. The German National Library in cooperation with other German library developed an archiving and exchange package format for digital objects—called Universal Object Format (UOF). UOF package consists of content files packed to one file (ZIP or TAR) and a metadata file created according to METS Schema. Among obligatory attributes of files are:

- ID,
- MIMEtype,
- Created,
- Size,
- Checksum,
- ChecksumType.

Detailed requirements are formulated in *Co-operative Development of a Long-Term Digital Information Archive, Frankfurt am Main* 2006 [93].

The packages are transformed in the archive. During ingest process SIP is transformed to AIP or to a part of AIP, or to several AIPs. AIP can be enriched by information concerning preservation activities undertaken by the archive. Archival information requested by a user is prepared by transformation of appropriate AIP or AIPS to DIP.

Archival packages can also be used to transfer information between repositories. DIP sent from one archive can be accepted in another as SIP. Sending AIP directly to another archive is also considered. In both cases some details concerning construction of packages must be agreed between archives. A detailed discussion of such problems concerning transfer of packages is presented in *Enriched Archival Information Packages* [94].

Project LOTAR—Metadata for Archival Package Workgroup [86] claims that their objective is "is to develop, publish and maintain a standard for Metadata for Archive Packages in a neutral form that can be read and reused independently of changes in the IT application environment originally used for creation".

E-ARK project—European Archival Records and Knowledge Preservation—is oriented towards big data. Authors claim: "Special attention is paid to the requirement of being able to handle very large data sets which have to be split into several parts because it is not possible to archive them as a single coherent unit [87]".

5.2 Metadata Encoding and Transmission Standard (METS)

METS standard is maintained by the Library of Congress [6]. It was created to be used mainly by libraries. It is an XML Schema which allows for:

- expressing the structure of a digital entity by identifying files which comprise the content of the entity,
- linking descriptive metadata and other metadata with digital content,
- wrapping digital content and associated metadata.

Document METS [95] have following sections:

1. METS header,
2. descriptive metadata,
3. administrative metadata,
4. file section,
5. structural map section,
6. structural link section,
7. behavior section.

The header contains information concerning the METS document itself (identifiers, dates, identifiers of other sections or elements for internal linking, agents).

Descriptive metadata can refer to the METS object as a whole or to any of its components. METS does not define own descriptive metadata. Many descriptive metadata standards can be used, e.g. Dublin Core, MARC, MODS, VRA, EAD; local standards can also be used. Multiple sections are allowed; they can use different standards. METS provides a means to link metadata to digital content or to other metadata, e.g. administrative or structural (`<mdRef>`). Links can be defined also to non descriptive metadata, e.g. to preservation metadata PREMIS. There is also wrapper element defined (`<mdWrap>`) which can contain either Base64 encoded metadata or XML encoded metadata. Metadata Reference element mdRef can point to external location.

Administrative section contains metadata which refer to the object as a whole, to its components or to original source material. The section is divided into four subsections that contain:

- technical metadata,
- intellectual property rights,
- analog/digital source metadata,
- provenance metadata.

Way to handle these metadata is similar to that of descriptive type. METS does not define own standard of these metadata types. Linking and referencing can be made in the same way.

File section acts as inventory of content files that comprise the digital object described in the document. There is a `<file>` element for each file; a `<fileGrp>` element allows files to be grouped. Element `<file>` has attributes such as:

- ID—identifier unique within the METS document,
- MIMETYPE,
- SEQ—sequence of this file relative to the others in its group `<fileGrp>`,
- SIZE,

- CREATED,
- CHECKSUM—the checksum value,
- CHECKSUMTYPE—the checksum type (it must be one from the list accepted in the current version of METS Schema),
- OWNERID,
- ADMID—contains the ID attribute value of the `<techMD>`, `<sourceMD>`, `<rightsMD>` or `<digiprovMD>` elements within the administration section pertaining to the file,
- DMDID—contains the ID attribute value of the `<dmdSec>` pertaining to the file,
- GROUPID—relation to other file groups,
- USE.

Structural map organizes files in a consistent structure, e.g. hierarchical. There can be more than one structural map, e.g. one can determine logical structure of a book (chapters, sections, etc.) and the other physical structure (sequence of pages). In general structural map can link files containing text, sound, still images, video and others. Structure is described as a tree of nested elements `<div>` (division); links by pointers `<fptr>` and `<mptr>`.

Structural map is obligatory in METS document.

A number of Example Documents is presented on METS page [96]. The examples concern various types of digital objects or different domains, e.g. Bibliographic Record, Image with Text, Image with Video, Video with Transcript, Sheet Music, Page Turners, Maps & Geographic. Size of METS metadata files varies from about 10 kB to 200 kB, exceeding 1 MB in one case.

5.3 XML Formatted Data Unit (XFDU)

XFDU standard was created by *Consultative Committee for Space Data Systems* (CCSDS), international organization founded by space agencies for the development of communication and data handling standards. Member agencies come from such countries as USA, Russia, China, Japan, United Kingdom, France, Germany, Italy, Canada and Brazil. Also European Space Agency ESA is the member of CCSDS. About 30 other agencies from various countries have the observer status.

Main results of CCSDS activities are published as *Recommendations*, e.g. Recommended Standards or Recommended Practices. Recommended standards are based on consensus of the Committee members. In case of XFDU packages three standards have special meaning:

1. *XML Formatted Data Unit (XFDU). Structure and Construction Rules. Recommended standard. CCSDS 661.0-B-1. Blue Book, September 2008. ISO equivalent 13527:2010* citeGP:xfdu;
2. *Producer-Archive Interface Methodology Abstract Standard (PAIMAS). Recommendation for Space System Practices. CCSDS 651.0-M-1. Magenta Book. May 2004. ISO equivalent 20652:2006* [97];

3. *Producer-Archive Interface Specification (PAIS). Recommended Standard. Blue Book. Issue 1. February 2014. CCSDS 651.1-B-1. ISO equivalent 20104:2015* [98].

As it is defined in those standards, standard 1 (XFDU) "is a technical Recommendation to use for the packaging of data and metadata, including software, into a single package (e.g., file, document or message) to facilitate information transfer and archiving." The purpose of Recommendation 2 "is to identify, define and provide structure to the relationships and interactions between an information Producer and an Archive". Recommended Standard 3 "is a technical Recommendation providing the abstract syntax and an XML implementation of descriptions of data to be sent to an archive." It includes "one concrete implementation for the packages based on the XML Formatted Data Unit (XFDU) standard".

XFDU packaging is based on three concepts:

- Package Interchange File,
- Manifest Document,
- XFDU (XML Formatted Data Unit).

Package Interchange File is a container, in which the special file called Manifest Document must be included and can be bundled a number of other files. In the Manifest Document relations among the files are described and locations of all the files within the Package Interchange File are indexed. The Manifest can also contain files (data and metadata files) and have references to external files. Term XFDU is referred to Package Interchange File together with external files and packages referenced from the Manifest file.

XFDU functionality contains a few features: two packaging techniques are indicated in the XFDU standard: single XML document and archive file, and the XML messaging form is provided in the future, e.g. XML-binary Optimized Packaging XOP [99].

There are three Metadata/data linkage options:

- inclusion in Manifest as base64 or XML,
- referenced directly as binary or XML,
- referenced or included as Data Object.

There is one assumption concerning the encoding and transformations: "The ability to allow/reverse multiple transformations on files".

The XFDU Manifest file can include:

- Package Header—contains metadata which apply to the whole package;
- Information Package Map—uses ContentUnit elements which contain pointers to data objects and to their metadata;
- Data Objects—a byte stream and data necessary to reverse the transformation and return to original format;
- Metadata Objects—metadata can be recorded separately for each item;
- Behavior Objects (mainly for future use).

Package designer can define own metadata model. There are also predefined metadata which support the OAIS information model.

The construction and use of XFDU standard is illustrated by a number of examples [85].

6 Preservation Metadata

6.1 PREMIS Standard

PREMIS achieved great significance among preservation metadata standards. The name PREMIS stands for *PREservation Metadata: Implementation Strategies*. Originally it was name of international working group established in 2003 to design metadata for use in digital preservation. This group released in 2005 report: *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group (Version 1.0)* [100]. Project Credo used version 2.2 of PREMIS Data Dictionary, released in 2012 (next version 3.0 was not complete at that time). All three versions of standard PREMIS are available at [7]. Description of standard PREMIS in this chapter is based on version 2.2 of Data Dictionary for Preservation Metadata [101].

Construction of metadata in PREMIS standard is based on *data model* and *semantic units*.

Data model defines five interrelated (*Entities*): *Intellectual Entity, Object, Event, Agent* and *Rights*. The Intellectual Entity is defined as "a set of content that is considered a single intellectual unit for purposes of management and description". Other entities have following definitions:

- Object (or Digital Object): a discrete unit of information in digital form.
- Event: an action that involves or impacts at least one Object or Agent associated with or known by the preservation repository.
- Agent: person, organization, or software program/system associated with Events in the life of an Object, or with Rights attached to an Object.
- Rights: assertions of one or more rights or permissions pertaining to an Object.

The PREMIS data dictionary defines *Semantic units*; each of them is mapped to one of the entities. (In some cases semantic units can have subunits, e.g. *identifier* can have *identifierType* and *identifierValue*, or can have *extensions* to allow the use of metadata based on external schema, e.g. *environment* has subunit *environmentExtension*).

The Data Dictionary v. 2.2 defines four entities: Object, Event, Agent and Rights in more detailed way.

6.1.1 Object Entity

Semantic unit Object aggregates information concerning digital object and describes characteristics important for preservation management. There is only one obligatory element *objectIdentifier*, which refers to all object types.

There are three types of Object:

- *Representation*: the set of files needed for a complete presentation of *Intellectual Entity*, e.g. image files of pages of a book and structural metadata;
- *File*: ordered sequence of bytes known by an operating system;
- *Bitstream*: a set of bits embedded within a file that has meaningful common properties for preservation purposes.

The above definition of bitstream differs from common usage. A detailed explanation is given in the Data Dictionary.

An Object can be linked to one or more *rightsStatements* and can take part in one or more *Events*. Following semantic units of Object are defined:

Listing 15 Semantic units of Object

```
1 objectIdentifier (M, R)
  subunits Type and Value
2 objectCategory (M, NR)
3 preservationLevel (O, R) [representation, file]
  subunits: Value, Role, Rationale and DateAssigned
4 significantProperties (O, R)
  subunits: Type, Value, Extension
5 objectCharacteristics (M, R) [file, bitstream]
  subunits: compositionLevel, Fixity
        (messageDigestAlgorithm, messageDigest,
        messageDigestOriginator), size, format
        (formatDesignation (formatName, formatVersion),
        formatRegistry (Name, Key, RoleR), formatNote),
        creatingApplication (creatingApplicationName,
        creatingApplicationVersion,
        dateCreatedByApplication,
        creatingApplicationExtension), inhibitors
        (Type, Target, Key),
        objectCharacteristicsExtension
6 originalName (O, NR) [representation, file]
7 storage (O, R) [file, bitstream]
  subunits: contentLocation (Type, Value), storageMedium
8 environment (O, R)
  subunits: environmentCharacteristic,
        environmentPurpose, environmentNote, dependency
        (Name, Identifier (Type, Value), software
        (swName, swVersion, swType, swOtherInformation,
        swDependency), hardware (hwName, hwType,
        hwOtherInformation), environmentExtension
9 signatureInformation (O, R) [file, bitstream]
  subunits: signature (signatureEncoding, signer,
        signatureMethod, signature\-Value,
        signatureValidationRules, signatureProperties,
        keyInformation), signatureInformationExtension
10 relationship (O, R)
   subunits: relationshipType, relationshipSubType,
         relatedObjectIdentification
         (relatedObjectIdentifierType,
         relatedObjectIdentifierValue,
         relatedObjectSequence),
         relatedEventIdentification
```

```
          (relatedEventIdentifierType,
          relatedEventIdentifierValue,
          relatedEventSequence)
11 linkingEventIdentifier (O, R)
   subunits: linkingEventIdentifierType,
          linkingEventIdentifierValue
12 linkingIntellectualEntityIdentifier (O, R)
   subunits: linkingIntellectualEntityIdentifierType,
          linkingIntellectualEntityIdentifierValue
13 linkingRightsStatementIdentifier (O, R)
   subunits: linkingRightsStatementIdentifierType,
          linkingRightsStatementIdentifierValue
```

Fixity is particularly important. Its subelements are: *messageDigestAlgorithm, messageDigest, messageDigestOriginator*. Typical preservation activity is validating checksum. Information about performing the test and the date should be recorded as *Event*; result of the test can be recorded as *eventOutcome*; in the *Object* should be recorded name of algorithm, e.g. *messageDigestAlgorithm* = MD5, *messageDigest* = 7c9b35da4f…and the agent as *messageDigestOriginator*.

Data Dictionary contains detailed information concerning particular elements.

6.1.2 Event Entity

The *Event* aggregates information on activities concerning one or more objects. Metadata concerning the events usually are stored outside the objects. The archive is not obliged to record all events, only important ones. Events modifying the object should always be recorded.

Elements: *eventIdentifier, eventType, eventDateTime* should always be used.

Event must be related to one or more objects and can be related to one or more agents.

Listing 16 Semantic units of *Event*

```
1 eventIdentifier (M, NR)
    eventIdentifierType (M, NR)
    eventIdentifierValue (M, NR)
2 eventType (M, NR)
3 eventDateTime (M, NR)
4 eventDetail (O, NR)
5 eventOutcomeInformation (O, R)
    eventOutcome (O, NR)
    eventOutcomeDetail (O, R)
      eventOutcomeDetailNote (O, NR)
      eventOutcomeDetailExtension (O, R)
6 linkingAgentIdentifier (O, R)
    linkingAgentIdentifierType (M, NR)
    linkingAgentIdentifierValue (M, NR)
    linkingAgentRole (O, R)
7 linkingObjectIdentifier (O, R)
    linkingObjectIdentifierType (M, NR)
    linkingObjectIdentifierValue (M, NR)
    linkingObjectRole (O, R)
```

6.1.3 Agent Entity

Semantic unit *Agent* aggregates information on attributes and characteristics of agents concerning the rights management and the management of preservation events in the archive. The only obligatory element is *agentIdentifier*. Agent can have rights and can grant rights—one or more. Agent can perform events, authorize. Agent can affect one or more objects by the events or based on (*rights statements*).

Listing 17 Semantic units of Agent

```
1 agentIdentifier (M, R)
    agentIdentifierType (M, NR)
    agentIdentifierValue (M, NR)
2 agentName (O, R)
3 agentType (O, NR)
4 agentNote (O, R)
5 agentExtension (O, R)
6 linkingEventIdentifier (O, R)
    linkingEventIdentifierType (M, NR)
    linkingEventIdentifierValue (M, NR)
7 linkingRightsStatementIdentifier (O, R)
    linkingRightsStatementIdentifierType (M, NR)
    linkingRightsStatementIdentifierValue (M, NR)
```

6.1.4 Examples of PREMIS Metadata

Data dictionary v. 1.0 contains 6 examples, showing in details how PREMIS metadata can be applied where objects have different structure:

- Microsoft Word document complete in one file,
- Electronic Theses and Dissertation (two files: a PDF and MP3),
- Newspaper complex object, Los Angeles Times (tar file that contains within a file produced by application QuarkXPress, which in turn contains links to EPS file),
- Web site (harvested Web site),
- Digital signature,
- Photograph (two files: a TIFF image and an XML file containing descriptive metadata).

For example, in case 1 Word file was ingested to archive (repository). Two digital objects were created: representation and the file. Five events were recorded: Ingest, Fixity check, Virus check, Object validation and Annotation validation.

Further examples can be found at [102].

6.1.5 Changes

PREMIS is under constant development. Procedures for requesting a change to the PREMIS Data Dictionary or to the associated XML Schema are defined at [103].

Among changes introduced to PREMIS Data Dictionary for Preservation Metadata version 3.0 are:

- Intellectual Entity became another category of Object, having the same semantic units as Representation.
- A new semantic unit *preservationLevelType* was defined to indicate preservation activities which might be applied to the object for given *preservationLevel*.
- A new semantic unit *agentVersion* was added to express the version of software *Agents*.

6.2 Besides PREMIS

The German National Library has presented in 2005 a schema for recording suitable technical metadata in the form of long-term preservation metadata for electronic resources (LMER) [104].

The project was inspired by earlier simple preservation metadata standard created by the National Library of New Zealand (2003) [105]. PREMIS did not exist at that time. The National Library of New Zealand took part in the works on PREMIS from the beginning, sharing experiences with working group.

Implementation of PREMIS can be a challenge for small or medium size libraries or archives. Some centers tried to design simpler sets of preservation metadata. An example of such approach can be found in North Carolina Preservation Metadata for Digital Objects (PMDO) [106]. System is simple—only 21 metadata elements. What is interesting, each element has mapping to PREMIS, so that institution using the PMDO can easily change it to PREMIS.

PREMIS can inspire integration with other standards. An interesting example of integrating other metadata standards with PREMIS is presented in [107]. The authors propose a model of provenance description based on integration of PREMIS OWL Ontology and ontology PROV-O.

7 Conclusion

Metadata play important role in long-term digital preservation. Metadata accompanying digital objects in current use, e.g. descriptive, technical and rights metadata, should be sent to archive together with the objects. It is recommended to gather, also from the beginning, provenance metadata concerning digital objects and send then to the archive with the objects. The Reference Model for an Open Archival Information System (OAIS) was developed to standardize digital preservation activities. Two metadata types play special role in long-term preservation: packaging metadata, applied to create packages SIP, AIP and DIP, and preservation metadata, which supports and documents process of preservation.

References

1. Digital Preservation Handbook Glossary. Digital Preservation Coalition http://www.handbook.dpconline.org/glossary. Access: 2016-10-20.
2. Digital Preservation Metadata. Angela Dappert, The British Library, Planets, PREMIS EC, Barcelona, March 2009 http://www.planets-project.eu/docs/presentations/Dappert_PreservationMetadata.pdf. Access: 2016-10-20.
3. What is Provenance. W3C http://www.w3.org/2005/Incubator/prov/wiki/What_Is_Provenance. Access: 2016-10-20.
4. Visual Resources Association http://www.vraweb.org/index.html. Access: 2016-10-20.
5. Directory Interchange Format DIF. Global Change Master Directory http://www.gcmd.nasa.gov/add/difguide/index.html. Access: 2016-10-20.
6. METS: Metadata Encoding and Transmission Standard http://www.loc.gov/standards/mets/. Access: 2016-10-20.
7. PREMIS Data Dictionary for Preservation Metadata http://www.loc.gov/standards/premis/. Access: 2016-10-20.
8. Simple Dublin Core https://www.dublincore.org/documents/1998/09/dces/. Access: 2016-10-20.
9. EBU Core metadata set for Radio archives, Tech 3293, version 1.0, 2001 http://www.tech.ebu.ch/publications/tech3293. Access: 2016-10-20.
10. European Broadcasting Union http://www.ebu.ch/home. Access: 2016-10-20.
11. Dublin Core Element Set to GCMD Directory Interchange Format DIF. Global Change Master Directory http://www.gcmd.nasa.gov/add/standards/dublin_to_dif.html. Access: 2016-10-20.
12. The Open Archives Initiative Protocol for Metadata Harvesting, v. 2.0, http://www.openarchives.org/OAI/openarchivesprotocol.html. Access: 2016-10-20.
13. Dublin Core Qualifiers http://www.dublincore.org/documents/2000/07/11/dcmes-qualifiers/. Access: 2016-10-20.
14. DCMI Metadata Terms http://www.dublincore.org/documents/dcmi-terms/. Access: 2016-10-20.
15. MARC Code List for Relators http://www.loc.gov/marc/relators/relaterm.html. Access: 2016-10-20.
16. Guidelines for implementing Dublin Core in XML http://www.dublincore.org/documents/dc-xml-guidelines/. Access: 2016-10-20.
17. XML Schema for Simple Dublin Core 2002 http://www.dublincore.org/schemas/xmls/simpledc2002/1212.xsd. Access: 2016-10-20.
18. XMLSchema for DC Terms http://www.dublincore.org/schemas/xmls/qdc/dcterms.xsd. Access: 2016-10-20.
19. arXiv.org Cornell University Library http://www.arxiv.org. Access: 2016-10-20.
20. Federacja Bibliotek Cyfrowych (FBC) http://www.fbc.pionier.net.pl. Access: 2016-10-20.
21. EBU Core Metadata Set (EBUCore). Specification 1.6. Tech 3293. Geneva, June 2015 http://www.tech.ebu.ch/docs/tech/tech3293.pdf. Access: 2016-10-20.
22. SMPTE – Society of Motion Picture and Television Engineers http://www.smpte.org. Access: 2016-10-20.
23. International Telecommunication Union http://www.itu.int/en/about/Pages/default.aspx. Access: 2016-10-20.
24. International Press Telecommunications Council (IPTC) http://www.iptc.org/. Access: 2016-10-20.
25. Metadata Guidance for the Transfer of Permanent Electronic Records. NARA Bulletin 2015-04, September 15, 2015 http://www.archives.gov/records-mgmt/bulletins/2015/2015-04.html. Access: 2016-10-20.
26. Sustainability of Digital Formats Planning for Library of Congress Collections http://www.digitalpreservation.gov/formats/. Access: 2016-10-20.

27. Revised Format Guidance for the Transfer of Permanent Electronic Records. NARA Bulletin 2014-04, February 4, 2014, http://www.archives.gov/records-mgmt/bulletins/2014/2014-04.html. Access: 2016-10-20.
28. List of Metadata Standards. Digital Curation Centre (DCC) http://www.dcc.ac.uk/resources/metadata-standards/list. Access: 2016-10-20.
29. List of metadata standards in Wikipedia http://www.en.wikipedia.org/wiki/Metadata_standard. Access: 2016-10-20.
30. Domain Metadata Standards. University of Central Florida Libraries http://www.guides.ucf.edu/metadata/domMetaStandards. Access: 2016-10-24.
31. Agricultural Metadata Element Set (AgMES) – AgMES 1.1 Namespace Specification http://www.aims.fao.org/standards/agmes/namespace-specification. Access: 2016-10-20.
32. AVM – Astronomy Visualization Metadata, International Virtual Observatory Alliance standard used to tag astronomical images http://www.virtualastronomy.org/avm_metadata.php. Access: 2016-10-20.
33. PROV provenance metadata standard PROV[89] designed for interchange of provenance information http://www.w3.org/2001/sw/wiki/PROV. Access: 2016-10-20.
34. VRA Core Schemas & Documentation. Official Website http://www.loc.gov/standards/vracore/. Access: 2016-10-24.
35. MODS: Metadata Object Description Schema http://www.loc.gov/standards/mods/. Access: 2016-10-20.
36. TEI: Text Encoding Initiative http://www.tei-c.org/index.xml. Access: 2016-10-20.
37. ONIX standards: ONIX for Books, ONIX for Subscription Products, Licencing Terms and Right Information http://www.editeur.org/8/ONIX/. Access: 2016-10-20.
38. STEP: Standard for the Exchange of Product model data http://www.steptools.com/library/standard/. Access: 2016-10-24. – also ISO 10303: Industrial automation systems and integration – Product data representation and exchange (has many parts).
39. Metadata for Long Term Preservation of Product Data. International Symposium on XML for the Long Haul: Issues in the Long-term Preservation of XML. Proceedings August 2, 2010 http://www.balisage.net/Proceedings/vol6/html/Lubell01/BalisageVol6-Lubell01.html. Access: 2016-10-20.
40. MARC to Dublin Core Crosswalk http://www.loc.gov/marc/marc2dc-2001.html. Access: 2016-10-20.
41. Dublin Core to MARC Crosswalk http://www.loc.gov/marc/dccross.html. Access: 2016-10-20.
42. Mapping of the Dublin Core Metadata Element Set to the CIDOC CRM, Martin Doerr http://www.cidoc-crm.org/docs/dc_to_crm_mapping.pdf. Access: 2016-10-24.
43. The CIDOC Conceptual Reference Model. International Council of Museums (ICOM) http://www.cidoc-crm.org/official_release_cidoc.html. Access: 2016-10-24.
44. Europeana Collections http://www.europeana.eu. Access: 2016-10-20.
45. Provide data in EDM. Europeana Pro http://www.pro.europeana.eu/page/provide-data-edm. Access: 2016-10-20.
46. New railroad map of the state of Maryland, Delaware, and the District of Columbia, available at http://www.loc.gov/item/98688491/; MODS record http://www.lccn.loc.gov/98688491/mods; Dublin Core record http://www.lccn.loc.gov/98688491/dc. Access: 2016-10-20.
47. IPTC Photo Metadata Standard (2014) http://www.iptc.org/standards/photo-metadata/iptc-standard/. Access: 2016-10-24.
48. IPTC Core and Extension Guidelines (July 2010) http://www.iptc.org/standards/photo-metadata/iptc-standard/. Access: 2016-10-24.
49. IPTC Core and Extension examples http://www.iptc.org/std/photometadata/documentation/GenericGuidelines/index.htm#!Documents/iptccoreandextensionexamples.htm. Access: 2016-10-24.
50. Museum of History of Photography http://www.mhf.krakow.pl. Access: 2016-10-24.
51. History Meeting House http://www.dsh.waw.pl. Access: 2016-10-24.

52. Exchangeable image file format for digital still cameras: Exif Version 2.31. Standard of the Camera & Imaging Products Association. CIPA DC-008-Translation-2016. http://www.cipa.jp/std/documents/e/DC-008-Translation-2016-E.pdf. Access: 2016-10-24.
53. Exchangeable image file format for digital still cameras: Exif Version 2.2. April, 2002. Technical Standardization Committee on AV & IT Storage Systems and Equipment. Japan Electronics and Information Technology Industries Association. http://www.exiv2.org/Exif2-2.PDF. Access: 2016-10-24.
54. Exif vocabulary workspace – RDF Schema http://www.w3.org/2003/12/exif/. Access: 2016-10-24.
55. MIX – NISO Metadata for Images in XML Schema. Technical Metadata for Digital Still Images Standard http://www.loc.gov/standards/mix/. Access: 2016-10-24.
56. American National Information Standards Organization http://www.niso.org/home/. Access: 2016-10-24.
57. TIFF. Revision 6.0. Final June 3, 1992. Adobe Developers Association http://www.partners.adobe.com/public/developer/en/tiff/TIFF6.pdf. Access: 2016-10-24.
58. XMP Specification. Adding Intelligence to Media. Adobe Systems Incorporated. September 2005 http://www.partners.adobe.com/public/developer/en/xmp/sdk/XMPspecification.pdf. Access: 2016-10-24.
59. Digital Negative (DNG) Specification, Version 1.4.0.0, June 2012 http://www.images.adobe.com/content/dam/Adobe/en/products/photoshop/pdfs/dng_spec_1.4.0.0.pdf. Access: 2016-10-24.
60. International Association of Sound and Audiovisual Archives (IASA) http://www.iasa-web.org. Access: 2016-10-24.
61. IASA Cataloguing Rules http://www.iasa-web.org/cataloguing-rules. Access: 2016-10-24.
62. IASA Technical Committee, Guidelines on the Production and Preservation of Digital Audio Objects, ed. by Kevin Bradley. Second edition 2009. (= Standards, Recommended Practices and Strategies, IASA-TC 04) http://www.iasa-web.org/tc04/audio-preservation. Access: 2016-10-24.
63. ISO/IEC 13818-3:1995 Information technology "Generic coding of moving pictures and associated audio information" Part 3: Audio: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=22412. Access: 2016-10-24.
64. EBU Classification Schemes and ontologies: www.ebu.ch/metadata/cs/ and www.ebu.ch/metadata/ontologies/. Access: 2016-10-20.
65. EBU Role Code Classification Scheme http://www.ebu.ch/metadata/cs/ebu_RoleCodeCS.xml. Access: 2016-10-20.
66. FIAF – Fédération Internationale des Archives du Film / International Federation of Film Archives http://www.fiafnet.org. Access: 2016-10-24.
67. IFTA – International Federation of Television Archives[42] http://www.fiatifta.org. Access: 2016-10-24.
68. FIAF Cataloguing Rules for Film Archives http://www.fiafnet.org/images/tinyUpload/E-Resources/Commission-And-PIP-Resources/CDC-resources/FIAF_Cat_Rules.pdf. Access: 2016-10-24.
69. The FIAF Moving Image Cataloging Manual. Written by Natasha Fairbairn, Maria Assunta Pimpinelli, Thelma Ross, April 2016 http://www.fiafnet.org/images/tinyUpload/E-Resources/Commission-And-PIP-Resources/CDC-resources/20160920%20Fiaf%20Manual-WEB.pdf. Acess: 2016-10-24.
70. European Standards Committee (CEN) Cinematographic Works Standard (CWS) http://www.filmstandards.org/fsc.index.php/Main_Page. Access: 2016-10-24.
71. DIN EN 15744 Film identification – Minimum set of metadata for cinematographic works http://www.en-standard.eu/din-en-15744-film-identification-minimum-set-of-metadata-for-cinematographic-works/. Access: 2016-10-24.
72. DIN EN 15907 Film identification – Enhancing interoperability of metadata – Element sets and structures http://www.en-standard.eu/din-en-15907-film-identification-enhancing-interoperability-of-metadata-element-sets-and-structures/. Access: 2016-10-24.

73. EN 15744 Film identification Minimum set of metadata for cinematographic works http://www.filmstandards.org/fsc/index.php/EN_15744. Access: 2016-10-24.
74. EN 15907 Film identification – Enhancing interoperability of metadata – Element sets and structures http://www.filmstandards.org/fsc/index.php/EN_15907. Access: 2016-10-24.
75. Cox, M., Mulder, E., Tadic, L., Descriptive Metadata for Television: An End-to-End Introduction. Focal Press 2013.
76. TV-Anytime News. Schemas and Classification Schemas http://www.tech.ebu.ch/tvanytime. Access: 2016-10-24.
77. PBCore – Public Broadcasting Metadata Dictionary http://pbcore.org. Access: 2016-10-24.
78. P/META Metadata Library. EBU Tech 3295. Specification 2.2. Geneva, September 2011 http://www.tech.ebu.ch/docs/tech/tech3295v2_2.pdf. Access: 2016-10-24.
79. SPTME Metadata Registry class 13-14 – managed by Society of Motion Picture & Television Engineers (SMPTE) Registration Authority http://www.smpte-ra.org/smpte-metadata-registry. Access: 2016-10-24.
80. EBU Archives Report 2010. Technical Report 006. Geneva, June 2010 http://www.tech.ebu.ch/docs/techreports/tr006.pdf. Access: 2016-10-20.
81. Open Digital Rights Language (ODRL) http://www.w3.org/community/odrl/. Access: 2016-10-24.
82. DDEX Standards http://www.ddex.net/ddex-standards. Access: 2016-10-24.
83. METS RightsDeclarationMD Extension Schema http://www.loc.gov/standards/rights/METSRights.xsd. Access: 2016-10-24.
84. Guidelines for using PREMIS with METS for exchange, 2008 http://www.loc.gov/standards/premis/guidelines-premismets.pdf. Access: 2016-10-24.
85. XML Formatted Data Unit (XFDU). Structure and Construction Rules. Recommended standard. CCSDS 661.0-B-1. Blue Book, September 2008. ISO equivalent 13527:2010 http://www.public.ccsds.org/Pubs/661x0b1.pdf. Access: 2016-10-24.
86. LOTAR: Long Term Archiving and Retrieval. Metadata for Archival Package Workgroup http://www.lotar-international.org/lotar-workgroups/metadata-for-archival-package.html. Access: 2016-10-24.
87. E-ARK. WP4 Archival Records Preservation. Work Packages. http://www.eark-project.com/about/work-packages/9-about/32-wp4intro. Access: 2016-10-24.
88. METS Implementation Registry http://www.loc.gov/standards/mets/mets-registry.html. Access: 2016-10-24.
89. METS registered profiles http://www.loc.gov/standards/mets/mets-registered-profiles.html. Access: 2016-10-24.
90. CCSDS Missions http://www.public.ccsds.org/implementations/missions.aspx. Access: 2016-10-24.
91. Project CASPAR – Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval – an EU Integrated Project http://www.cordis.europa.eu/project/rcn/92920_en.html. Access: 2016-10-24.
92. Project SAFE – Standard Archive Format for Europe http://www.earth.esa.int/SAFE/. Access: 2016-10-24.
93. Universal Object Format An archiving and exchange format for digital objects. Project kopal Co-operative Development of a Long-Term Digital Information Archiv. By Dipl.-Inform. Tobias Steinke Frankfurt am Main 2006 http://www.kopal.langzeitarchivierung.de/index_objektspezifikation.php.en. Access: 2016-10-24.
94. Caplan, P., Repository to Repository Transfer of Enriched Archival Information Packages. D-Lib Magazine, November/December 2008, vol. 14, number 11/12. http://www.dlib.org/dlib/november08/caplan/11caplan.html. Access: 2016-10-20.
95. METS. Metadata Encoding and Transmission Standard: Primer and Reference Manual. v. 1.6, 2010, Digital Library Federation http://www.loc.gov/standards/mets/mets-schemadocs.html. Access: 2016-10-24.
96. METS Example Documents http://www.loc.gov/standards/mets/mets-examples.html. Access: 2016-10-24.

97. Producer-Archive Interface Methodology Abstract Standard (PAIMAS). Recommendation for Space System Practices. CCSDS 651.0-M-1. Magenta Book. May 2004. ISO equivalent 20652:2006 http://www.public.ccsds.org/Pubs/651x0m1.pdf. Access: 2016-10-24.
98. Producer-Archive Interface Specification (PAIS). Recommended Standard. Blue Book. Issue 1. February 2014. CCSDS 651.1-B-1. ISO equivalent 20104:2015 http://www.public.ccsds.org/Pubs/651x1b1.pdf. Access: 2016-10-24.
99. XML-binary Optimized Packaging, W3C Recommendation, 25 January 2005 http://www.w3.org/TR/xop10/. Access: 2016-10-24.
100. Preservation Metadata: Implementation Strategies http://en.wikipedia.org/wiki/Preservation_Metadata:_Implementation_Strategies. Access: 2016-10-24.
101. PREMIS Data Dictionary for Preservation Metadata. v. 2.2, July 2012, PREMIS Editorial Committee http://www.loc.gov/standards/premis/v2/premis-2-2.pdf. Access: 2016-10-24.
102. PREMIS examples http://www.loc.gov/standards/premis/examples.html. Access: 2016-10-24.
103. Procedures for requesting a change to the PREMIS Data Dictionary or to the associated XML Schema http://www.loc.gov/standards/premis/revision-process.html. Access: 2016-10-24.
104. Long-term preservation Metadata for Electronic Resources (v. 1.2) http://www.dnb.de/EN/Standardisierung/LMER/lmer_node.html. Access: 2016-10-24.
105. National Library of New Zealand. Preservation Metadata. Metadata Standards Framework Metadata Implementation Schema. July 2003 http://www.digitalpreservation.natlib.govt.nz/assets/Uploads/nlnz-data-model-final.pdf. Access: 2016-10-24.
106. NC Preservation Metadata for Digital Objects. 2013 EDITION BY LISA GREGORY, KATHLEEN KENNEY, AMY RUDERSDORF http://www.digitalpreservation.ncdcr.gov/pmdo2013final.pdf. Access: 2016-10-24.
107. Provenance Description of Metadata using PROV with PREMIS for Long-term Use of Metadata, Chunqiu Li, Shigeo Sugimoto. DC-2014 in Austin, Texas, 8-11 October 2014 http://www.dcpapers.dublincore.org/pubs/article/view/3709. Access: 2016-10-24.

http://www.springer.com/978-3-319-51800-8

Digital Preservation: Putting It to Work
Traczyk, T.; Ogryczak, W.; Pałka, P.; Śliwiński, T. (Eds.)
2017, XI, 158 p. 13 illus., Hardcover
ISBN: 978-3-319-51800-8