DANUBIUS-PP

Preparatory Phase for the pan-European Research Infrastructure DANUBIUS–RI "The International Centre for advanced studies on river-sea systems"

Metadata description and formalizing documentation procedures

Deliverable 7.3

European Commission

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 739562

Project Full title	Preparatory Phase for the pan-European Research Infrastructure DANUBIUS–RI "The International Centre for advanced studies on river-sea systems"
Project Acronym	DANUBIUS-PP
Grant Agreement No.	739562
Coordinator	Dr. Adrian Stanica
Project start date and duration	1st December 2016, 36 months
Project website	www.danubius-pp.eu

Deliverable Nr.	7.3	**Deliverable Date**	M12
Work Package No.	7		
Work Package Title	Metadata description and formalizing documentation procedures		
Responsible	INSB		
Authors & Institutes Acronyms	ROMANIA ***INSB*** Alexandru Amarioarei; Mihaela Paun; Andrei Paun; Manuela Sidoroff; Iris Tusa; Maria Paraschiv UK ***PML*** Steve Groom; Victor Martinez GERMANY ***BFG*** Stephan Dietrich		
Status:		**Final (F)**	●
		Draft (D)	

	Revised draft (RV)	
Dissemination level:	Public (PU)	
	Restricted to other program participants (PP)	
	Restricted to a group specified by the consortium (RE)	
	Confidential, only for members of the consortium (CO)	●

Version	Date	Authors	Changes
0.1	June 2017	Mihaela Paun	Initial draft
0.2	June 2017	Andrei Paun	Second Draft
0.3	August 2017	Alexandru Amarioarei	Third Draft
0.4	September 2017	Alexandru Amarioarei	Fourth Draft
0.5	November 2017	Alexandru Amarioarei	Fifth Draft

Executive summary / abstract

The DANUBIUS research infrastructure will implement one or several components for data delivery, services and applications intended to provide direct access to the necessary tools, policies, standards and communities that comprise DANUBIUS-RI. The metadata will provide the information needed for these components to connect end-users with available data, services and applications.

The types of metadata vary from one environment to another and this deliverable will describe various categories of metadata, their attributes, functions and standards by DANUBIUS-RI area available internationally. The following sections will present the metadata description and the documentation procedures required, however the decision for which metadata standards DANUBIUS-RI will implement will be reached by the consortium and be presented in the Deliverables 7.5 and 7.7

Table of Contents

List of Figures

List of Tables

1. Introduction

The types of digital data or digital information that Danubius-RI will store will include:

- **datasets:** measurements of environmental variables obtained in the field, lab, or in simulations
- **reports:** discussion documents, synthesis reports, project reports, comments, literature reviews, user manuals, standard operating procedures, etc.
- **software for analyzing environmental research data:** software codes, software libraries, simulation models, etc.
- **images:** photographs, simulated images, remote sensing images, aerial photographs, etc.
- **cartographical output or GIS output:** GIS map, datasets with geographical objects and geographical coordinates

This digital information can be either publicly available and be downloaded through a web system. It can also get available via an authentication procedure. Then only certain files can be downloaded.

The past few years have seen increased attention to national and international policies for data archiving and sharing. One of the principal reasons includes the proliferation of digital data and a growing interest in research data and supplemental information as a part of the framework for scholarly communication. Key objectives cover not only preservation of scientific research data, but also the accessibility to verify research output and support the reuse and repurposing of data.

Metadata is clearly remarked in these undertakings and play a critical part to the success of any data repository or archiving initiative. Metadata is defined by the *New Merriam-Webster Dictionary*[1] as "data that provides information about other data" or simply put "data about data". The *National Information Standards Organization* (NISO[2]) defines metadata as "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource". The following three short statements about metadata in general can be viewed as "projections" of a three-dimensional definition of metadata:

- *Syntactical projection (representation-oriented)* - Metadata is physical representations of metainformation (metaknowledge) - as data are representations of information (knowledge).
- *Semantical projection (contents-oriented)* - Metadata provides information on data - and about processes of producing and using data.
- *Pragmatical projection (purpose-oriented)* - Metadata is data which are needed for proper production and use of the data they inform about.

Metadata allows users to locate and evaluate data without each person having to discover it again with every use. Its basic elements are a structured format and a controlled vocabulary,

[1] https://www.merriam-webster.com/dictionary/metadata

[2] http://www.niso.org

which together allow for a precise and comprehensible description of content, location, and value.

Metadata aspects can also be treated in several dimensions (Figure 1), by providing answers to questions one may have about their data: why? (e.g. was there data created? were values missing?) who? (e.g. created the data? manages the data?) what? (e.g. is the data content? source data was used?) when? (e.g. was the data created? is the time period of the content?) where? (e.g. is the study area? can the data be accessed?) how? (e.g. was the data created? are the data distributed?).

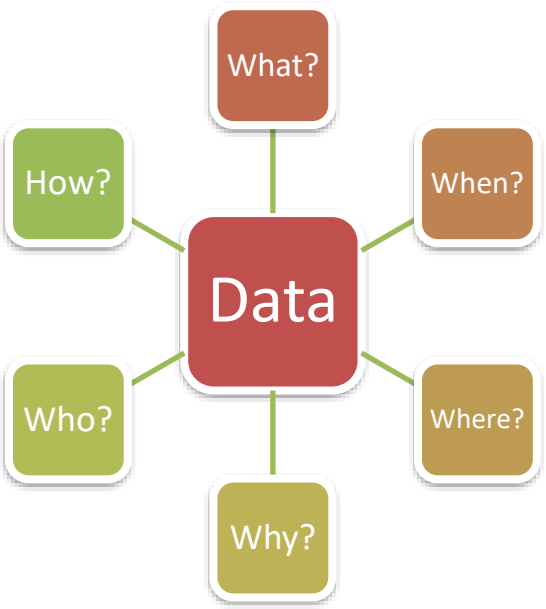

Figure 1. Dimensions of data

In addition to its use for locating and retrieving records and information consistently and accurately, metadata has a direct relationship to data management functions. It facilitates data sharing, ensures the authenticity of the content, describes the content, and facilitates implementation of retention and disposition. For electronic records management, metadata is critical to ensuring meaning, manageability, and preservation of electronic records.

The DANUBIUS-RI data infrastructure will be developed to improve data sharing and use among different sectors of policy makers, academia and the private sector, thus can be viewed as a base or structure of practices and relationships between data producers and users (Figure 2).

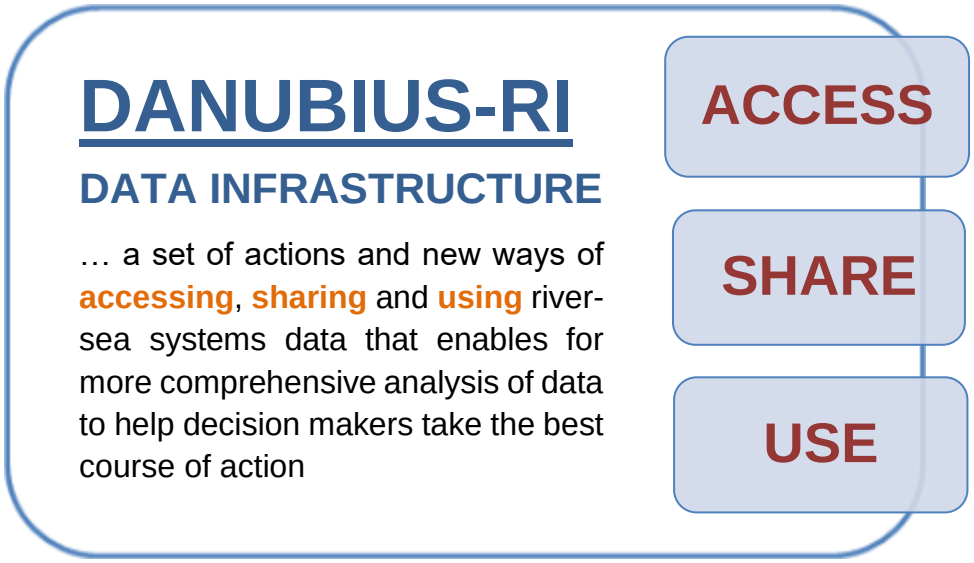

Figure 2. Data infrastructure diagram

The DANUBIUS research infrastructure will implement one or several components for data delivery, services and applications intended to provide direct access to the necessary tools, policies, standards and communities that comprise DANUBIUS-RI. The metadata will provide the information needed for these components to connect end-users with available data, services and applications (Figure 3).

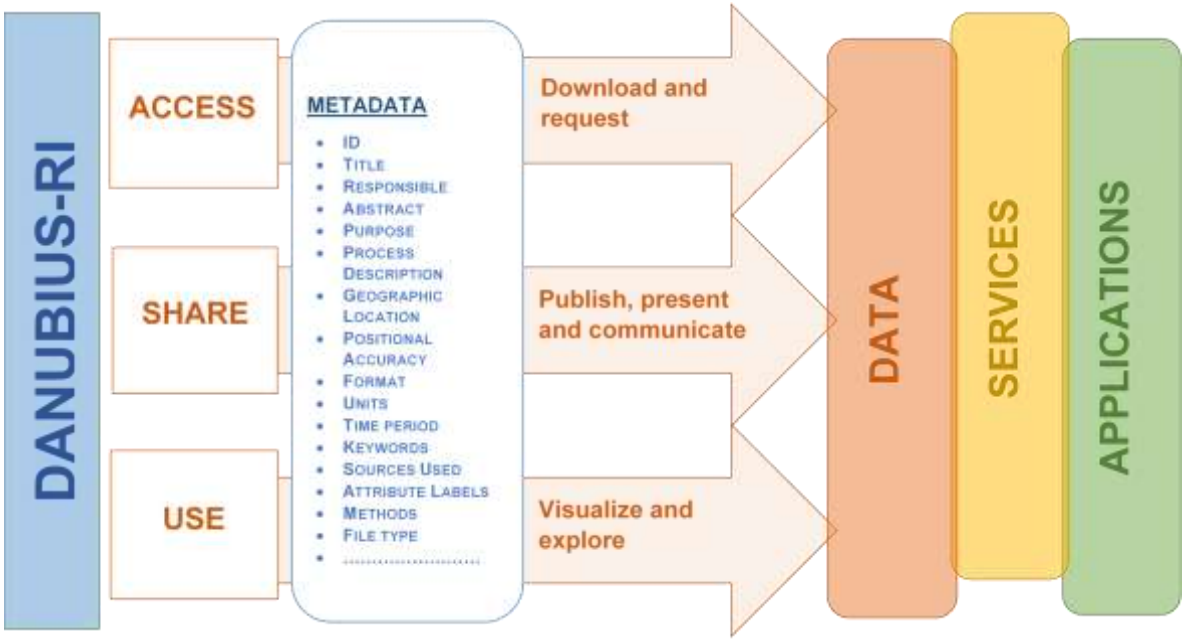

Figure 3. How metadata supports DANUBIUS-RI

This report is structured as follows:

- Section 2 describes the main types and functions of metadata
- Section 3 provides an overview of the descriptive metadata including the Simple Dublin Core metadata terms

- Section 0 discusses the second type of metadata in detail, namely administrative metadata
- Section 5 presents an overview of the metadata standards categorized both by discipline and subject areas
- Section 6 describes some of the metadata quality control metrics and their use

2. Types and functions of metadata

Though the types of metadata vary from one environment to another, it is helpful to separate metadata into distinct categories—descriptive, administrative (preservation, technical, rights) and use metadata—that reflect key aspects of metadata functionality (see Figure 4).

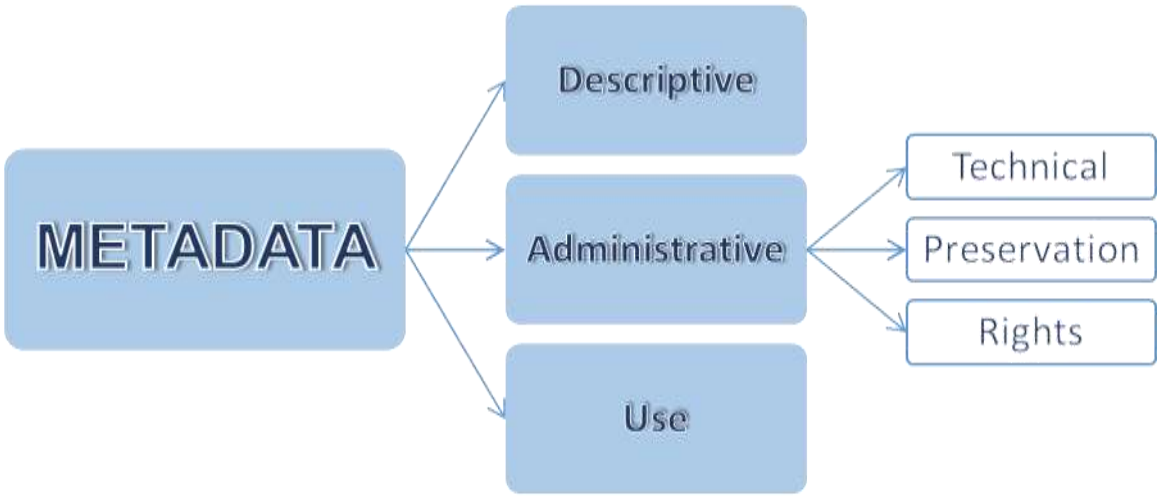

Figure 4. Types of metadata

Table 1 defines the metadata categories and gives examples of common functions that each might perform in a digital information system.

Table 1. Categories of metadata

Category / *Subcategory*	**Definition**	**Example**
Descriptive	*Metadata used to identify, authenticate, and describe collections and related trusted information resources*	- Metadata generated by original creator and system - Submission-information package - Cataloging records - Finding aids - Version control - Specialized indexes - Curatorial information - Linked relationships among resources - Descriptions, annotations, and emendations by creators and other users
Administrative	*Metadata used in managing and administering collections and information resources*	- Acquisition and appraisal information - Rights and reproduction tracking - Documentation of legal, cultural, and community-access requirements and protocols - Location information - Selection criteria for digitization - Digital repatriation documentation
Preservation	*Metadata related to the preservation management of collections and information resources*	- Documentation of physical condition of resources - Documentation of actions taken to preserve physical and digital versions of resources (e.g., data refreshing and migration) - Documentation of any changes occurring during digitization or preservation
Technical	*Metadata related to how a system functions or metadata behaves*	- Hardware and software documentation - System-generated procedural information (e.g., routing and event metadata) - Technical digitization information (e.g., formats, compression ratios, scaling routines) - Tracking of system-response times - Authentication and security data (e.g., encryption keys, passwords)
Use	*Metadata related to the level and type of use of collections and information resources*	- Circulation records - Physical and digital exhibition records - Use and user tracking - Content reuse and multiversioning information - Search logs - Rights metadata

Descriptive and *administrative metadata* are the most used types of metadata and we will focus on these two categories in sections 3 and 0. The *use metadata* refers to the use of data and is a fairly new development even if collecting data about the use of resources is not new as in the digital era web servers have software that collect logs about user activities (e.g. Google Analytics platform[3] or Coursera MOOC[4] dashboard data) and before that, libraries collected data about books cheked out and museums collected data about foot traffic through galleries.

In addition to its different types and functions, metadata exhibits many different characteristics. Table 2 presents some key characteristics of metadata, with examples. Metadata creation and management have become a complex mixture of manual and automatic processes and layers created by many different functions and individuals at different points during the life cycle of an information object. Effective and efficient metadata management is essential to ensure that the digital resources are trustworthy. Furthermore, the large volume of metadata that can potentially accumulate throughout the life of a resource is subject to a summarization and disposition regime.

Table 2. Attributes of metadata[5]

Attribute	Characteristics	Example
Source of metadata	Internal metadata generated by the creating agent for an information object at the time ot its first creation or digitalization Metadata intrinsic to an item or work	- File names and header information - Directory structures - File format and compression scheme - A title or inscription added to an artwork by its creator - A title or subtitle on the title page of a manuscript or printed book
	External metadata relating to an original item or information object; this is generated after the object has first been created or digitized, often by another person than the original creator	- URLs, URIs, PURLs, and other digital statements of provenance and online "location" - "Tracked" changes - Registrarial and cataloging records - Rights and other legal information
Method of metadata creation	Automatic creation, capture, or inferencing of metadata	- Keyword indexes - User-transaction logs - Audit trails
	Manual creation of metadata by information specialists	Descriptions of documentary interrelationships and intradocument relationships

[3] https://analytics.google.com/

[4] https://blog.coursera.org/whats-new-on-coursera-dashboard-and-course-home/

[5] https://www.getty.edu/publications/intrometadata/

Attribute	Characteristics	Example
	Manual or automatic creation of metadata during digitization processes	
	Individual user-contributed or crowd-sourced metadata	
Nature of metadata	Nonexpert metadata created by persons who are not subject or community specialists or information professionals	- Title HTML tags and meta tags created for a personal web page - Personal filing systems - Folksonomies
	Expert metadata created by subject or community specialists and/or information professionals, often not the original creator of the information object	- Specialized subject headings - Bibliographic records - Archival finding aids - Catalog entries for museum objects - Ad hoc metadata created by subject experts (e.g., tags added to an information object or catalog record by subject experts)
Structure	Structured metadata that conforms to a predictable standardized or proprietary structure	MARC, BIBFRAME, TEI, EAD, LIDO, local database formats
	Unstructured metadata that does not conform to a predictable structure	Unstructured note fields and other free-text annotations
Status	Static metadata that does not or should not change once it has been created	Technical information such as the date(s) of creation and modification of an information object, how it was created, file size
	Dynamic metadata that may change with use, manipulation, or preservation of an information object	- Directory structure - User-transaction logs
	Long-term metadata necessary to ensure that the information object continues to be accessible and usable	- Technical format and processing information - Rights information - Preservation management documentation
	Short-term metadata, mainly of a transactional nature	Interim location information
	Legacy metadata	Metadata created using an earlier system of metadata scheme
Semantics	Controlled metadata that conforms to a standardized vocabulary or authority form	- LCSH, LCNAF, AAT, ULAN, TGN, IA - AACR, RDA, DACS, CCO[6]

[6] LCSH – Library of Congress Semantic Head; LCNAF – Library of Congress Name Authority File; AAT – Getty Art and Architecture Thesaurus; ULAN – Getty Union List of Artist Names; TGN – Getty

Attribute	Characteristics	Example
	and that follows standard content (i.e., cataloging) rules	
	Uncontrolled metadata that does not conform to any standardized vocabulary or authority form	Free-text notes User-created tags
Level	Collection-level or group-level metadata relating to collections or groupings of original items and/or information objects	- Collection- or group-level record (e.g., a bibliographic record for a group or collection of items; a finding aid for an intact archival collection) - Series- or group-level information in a bibliographic record, finding aid, or museum collection record
	Item-level or within-item-level metadata relating to individual items and/or information objects, often contained within collections	- Catalog records for individual bibliographic items or unique cultural objects - Transcribed image captions and dates - "Tombstone" information for works of art and material culture - Format information

Some of the primary functions of metadata are presented in Table 3.

Table 3. Primary functions of metadata

Function	Description
Creation, multiversioning, reuse, and recontextualization of information objects	Objects enter a digital information system by being created digitally or by being converted into a digital format. Multiple versions of the same object may be created for preservation, research, exhibition, dissemination, or even product--development purposes. Some administrative and descriptive metadata may and indeed should be included by the creator or digitizer, especially if reuse is envisaged, such as in a digital asset management system.
Organization and description	A primary function of metadata is the description and ordering of original objects or items in a repository or collection as well as of the information objects relating to the originals. Information objects are automatically or manually organized into the structure of the digital information system and may include descriptions generated by the original creator. Additional metadata may be created by information professionals through registration, cataloging, and indexing

Thesaurus of Geographic Names; IA – Getty Iconography Authority; AACR – Anglo-American Cataloguing Rules; RDA – Resource Description and Access; DACS - Destination Addressing Control System; CCO – Cataloguing Cultural Objects

Function	Description
	processes, or by others via folksonomies and other forms of user-contributed metadata.
Validation	Users scrutinize metadata and other aspects of retrieved resources in order to ascertain the authoritativeness and trustworthiness of those resources.
Search and retreival	Good descriptive metadata is essential to users' ability to find and retrieve relevant metadata and information objects. Information objects—both those that are locally stored and virtually distributed—are subject to search and retrieval by users, and information systems create and maintain metadata that tracks retrieval algorithms, user transactions, and system effectiveness in storage and retrieval.
Utilization and preservation	In the digital realm, information objects may be subject to many different kinds of uses throughout their lives, during which they may also be reproduced and modified. Metadata related to user annotations, rights tracking, and version control may be created. Digital objects, especially those that are born digital, also need to be subject to a continuous preservation regime and undergo such processes as refreshing, migration, and integrity checking to ensure their continued availability and to document any changes that might have occurred to the information object during preservation processes.
Disposition	Metadata is a key component in documenting the disposition (e.g., accessioning, deaccessioning) of original objects and items in a repository as well as of the information objects relating to those originals. Information objects that are inactive or no longer necessary may be discarded.

A metadata *scheme or schema* is a minimum set of metadata that is well understood and used by everyone. It should be unambiguous. Metadata schemes for datasets are enabling original data in the science and social science fields to be shared. A metadata scheme provides the following facilities:

- A standardized way to locate network-accessible material.
- A consistent descriptive framework to store properties of the material.
- Facilitates queries that are more precise.
- A degree of fuzziness for queries (near, like, similar to).
- Groups objects into sets.
- Provides an order by which objects can be ranked.
- Access control.
- Commercial value and business logic support.
- Conversion and reuse for other purposes.
- Industrial strength workflow automation

3. Descriptive metadata

Descriptive metadata is typically used for discovery and identification, as information to search and locate an object or a resource. It can include elements such as title, author, subjects, keywords, publisher. A primary use of descriptive metadata records is for resource discovery but descriptive metadata may also be useful for maintenance of a resource.

One of the most used descriptive metadata schema is the Dublin Core Metadata, whose original objective was to define a set of elements that could be used by authors to describe their own web resources. The Dublin Core was developed to be simple and concise, and to describe web-based documents. Table 4 presents the initial set of 15 metadata elements of the Simple Dublin Core[7].

Table 4. The Simple Dublin Core matadata

Element	Description	Additional Information
Contributor	An entity responsible for making contributions to the resource.	Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.
Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.	Spatial topic and spatial applicability may be a named place or a location specified by its geographic coordinates. Temporal topic may be a named period, date, or date range. A jurisdiction may be a named administrative entity or a geographic place to which the resource applies. Recommended best practice is to use a controlled vocabulary such as the Thesaurus of Geographic Names [TGN]. Where appropriate, named places or time periods can be used in preference to numeric identifiers such as sets of coordinates or date ranges.
Creator	An entity primarily responsible for making the resource.	Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.
Date	A point or period of time associated with an event in the lifecycle of the resource.	Date may be used to express temporal information at any level of granularity. Recommended best practice is to use an encoding scheme, such as the

[7] http://www.niso.org/apps/group_public/download.php/10256/Z39-85-2012_dublin_core.pdf

Element	Description	Additional Information
		W3CDTF profile of ISO 8601 [W3CDTF].
Description	An account of the resource.	Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.
Format	The file format, physical medium, or dimensions of the resource.	Examples of dimensions include size and duration. Recommended best practice is to use a controlled vocabulary such as the list of Internet Media Types [MIME].
Identifier	An unambiguous reference to the resource within a given context.	Recommended best practice is to identify the resource by means of a string conforming to a formal identification system.
Language	A language of the resource.	Recommended best practice is to use a controlled vocabulary such as RFC 4646 [RFC4646].
Publisher	An entity responsible for making the resource available.	Examples of a Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.
Relation	A related resource.	Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system.
Rights	Information about rights held in and over the resource.	Typically, rights information includes a statement about various property rights associated with the resource, including intellectual property rights.
Source	A related resource from which the described resource is derived.	The described resource may be derived from the related resource in whole or in part. Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system.
Subject	The topic of the resource.	Typically, the subject will be represented using keywords, key phrases, or classification codes. Recommended best practice is to use a controlled vocabulary.
Title	A name given to the resource.	Typically, a Title will be a name by which the resource is formally known.

Element	Description	Additional Information
Type	The nature or genre of the resource.	Recommended best practice is to use a controlled vocabulary such as the DCMI Type Vocabulary [DCMITYPE]. To describe the file format, physical medium, or dimensions of the resource, use the Format element.

The initial set of metadata elements functioned in the context of three principles:

- *One-to-One* principle: the metadata record should only describe one manifestation or version of a resource
- *Dumb-Down* principle: the user should be able to look at the information in a metadata field with refinements or qualifications and still be able to make sense of the information if the refinements or qualifications were stripped away
- *Appropriate Value* principle: the metadata producer can never assume that their metadata would only be seen by a certain audience or in a certain context, so metadata should always be produced so that it would be understandable by any user in any context

Due to its simplicity and as response to new technological advances, the Dublin Core Metadata Element Set (DCMES[8]) has changed and developed significantly over the years, additional elements and refinements being added (Qualified Dublin Core) along with changes in the last two principles. To allow a complete and standardized description of a resource in Dublin Core, four terms that serve as building blocks have been proposed (according to *DCMI_MediaWiki User Guide*[9]):

- *Properties* - „core attributes of resources" that allow „uniform structured resource description"
- *Classes* - ways of grouping resources that have properties in common
- *Datatypes* - rules that govern how the information in certain properties are structured; can be used in properties such as dates, type, and format
- *Vocabulary Encoding Scheme* - vocabularies whose terms should be used to structure the information in properties such as creator, contributor, and subject

The following table presents the additional DCMI terms:

[8] http://dublincore.org/specifications/

[9] http://wiki.dublincore.org/index.php/User_Guide

Table 5. Additional DCMI terms[10]

Element	Refinement	Description
abstract	description	A summary of the resource.
accessRights	rights	Information about who can access the resource or an indication of its security status.
accrualMethod		The method by which items are added to a collection.
accrualPeriodicity		The frequency with which items are added to a collection.
accrualPolicy		The policy governing the addition of items to a collection.
alternative	title	An alternative name for the resource.
audience		A class of entity for whom the resource is intended or useful.
available	date	Date (often a range) that the resource became or will become available.
bibliographicCitation	identifier	A bibliographic reference for the resource.
conformsTo	relation	An established standard to which the described resource conforms.
created	date	Date of creation of the resource.
dateAccepted	date	Date of acceptance of the resource.
dateCopyrighted	date	Date of copyright.
dateSubmitted	date	Date of submission of the resource.
educationLevel	audience	A class of entity, defined in terms of progression through an educational or training context, for which the described resource is intended.
extent	format	The size or duration of the resource.
hasFormat	relation	A related resource that is substantially the same as the pre-existing described resource, but in another format.
hasPart	relation	A related resource that is included either physically or logically in the described resource.
hasVersion	relation	A related resource that is a version, edition, or adaptation of the described resource.

[10] http://dublincore.org/documents/dcmes-qualifiers/

Element	Refinement	Description
instructionalMethod		A process, used to engender knowledge, attitudes and skills, that the described resource is designed to support.
isFormatOf	relation	A related resource that is substantially the same as the described resource, but in another format.
isPartOf	relation	A related resource in which the described resource is physically or logically included.
isReferencedBy	relation	A related resource that references, cites, or otherwise points to the described resource.
isReplacedBy	relation	A related resource that supplants, displaces, or supersedes the described resource.
isRequiredBy	relation	A related resource that requires the described resource to support its function, delivery, or coherence.
issued	date	Date of formal issuance (e.g., publication) of the resource.
isVersionOf	relation	A related resource of which the described resource is a version, edition, or adaptation.
license	rights	A legal document giving official permission to do something with the resource.
mediator	audience	An entity that mediates access to the resource and for whom the resource is intended or useful.
medium	format	The material or physical carrier of the resource.
modified	date	Date on which the resource was changed.
provenance		A statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity, and interpretation.
references	relation	A related resource that is referenced, cited, or otherwise pointed to by the described resource.
replaces	relation	A related resource that is supplanted, displaced, or superseded by the described resource.
requires	relation	A related resource that is required by the described resource to support its function, delivery, or coherence.
rightsHolder	rights	A person or organization owning or managing rights over the resource.
spatial	coverage	Spatial characteristics of the resource.
tableOfContents	description	A list of subunits of the resource.

Element	Refinement	Description
temporal	coverage	Temporal characteristics of the resource.
valid	date	Date (often a range) of validity of a resource.

4. Administrative metadata

Administrative metadata provides information about the preservation, rights and access, and the technical issues needed to help manage a resource. It can answer questions such as when and how the resource was created, what file type or other technical elements it contains, and who will have access to it. A lot of it is interspersed and overlapping with and within the descriptive metadata. NISO[11] divides administrative metadata into three subcategories based on their common function and scope:

- *Technical metadata* - provides information about how a system functions, or system-level details about resources
- *Preservation metadata* - provides information to support the processes involved in ensuring that a resource continues to exist over time
- *Rights metadata* - provides information that may be used to control who gets access to a resource, under what conditions, and what they can do with it

4.1. Technical metadata

Technical metadata is used for recording the technical attributes of digital objects or datasets which helps a researcher and/or a repository manage digital objects over time. Some of the technical metadata include: the hardware and software used to acquire the digital object, file formats for master and derivatives, resolutions, color profiles, storage and location.

a) Textual data

For textual based digital objects, a recommended XML Schema designed for expressing technical metadata is textMD[12]. The elements, their description and related attributes are presented in Table 6.

[11] http://www.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf

[12] https://www.loc.gov/standards/textMD/

Table 6. TextMD Elements and attributes

	Element	Description
Top-Level Elements	encoding	Technical aspects of the text generation, whether analog-to-digital or born digital.
	character_info	Information regarding the encoding of characterswithin the file, including the standardized name of the character set, the byte order, the character size, and the line break mechanism.
	language	Language(s) used in work. Use ISO 639-2 codes, which are enumerated in the schema as valid text values.
	alt_language	A language code/description forthe text other than ISO 639-2. The alt_language elementhas a single attribute, authority, which may be used torecord the source of the language code (e.g., Ethnologue).
	font_script	The default font or script of the item.
	markup_basis	The metalanguage used to create the markup language, such as SGML, XML, GML, etc.
	markup_language	Markup language employed on the text (i.e., the specific schema or dtd). May be a URI for schema or dtd, but not mandatory.
	processingNote	Any general note about the processing of the file not covered elsewhere.
	printRequirements	Any special requirements for printing the item.
	viewingRequirements	Any special hardware or software requirements for viewing the item.
	textNote	Any general note on material not covered elsewhere.
	pageOrder	The natural (language-specific) page turning order of the text (left-to-right for Latin-based script, right-to-left for Arabic, Hebrew, etc.) independent of how it is represented in the METS file.
	pageSequence	The arrangement of the page-level divs in the METS file. That is, does the first div contain the first page a user would naturally read based on the language-specific direction of the text (the beginning of the content) or the last page the user would naturally read (the end of the content)? Enumerated values are ‘reading-order’ and ‘inverse-reading-order’.

	Element	Description
Low-Level Elements	byte_order	Byte order, primarily useful for cases where it's not clear just by specifying an IANA character set. Uses enumerated values of 'big,' 'little,' and 'middle' endian.
	byte_size	The size of an individual byte within theexpressed as a number of bits (as integer). This does not necessarilyequal the character size, as a character may have more thanone, or a variable number of bytes per character.
	character_size	The size of an individual character within thecharacter set as a number of bytes of the size expressed in thebyte_size. In the case of variable encodings, such as UTF-8for Unicode, the character_size element should state "variable"and also identify the specific variable character set encoding inthe encoding attribute.
	charset	The character set employed by the text. Controlled vocabusing IANA names for character sets.
	encoding_agent	Person who transcribed text from the original medium toanother. For example, in the case of an oral history transcript or a transcription of astone rubbing.
	encoding_platform	Hardware platform on which document was original produced, including specific computer type and any imaging equipment used for OCR.
	encoding_software	Type of software used in producing text, including OCR, word processing, text editor, etc.
	linebreak	How line breaks are represented in currentfile (which may differ from how they were originally encoded). Either carriage return, line feed, or carriage return/line feed.
Attributes	authority	A string used to record the source of the non-ISO 639-2 language code (e.g., Ethnologue).
	encoding	Used to identify a specific variable character set (as a string), such as UTF-8.
	linebreak	Used to indicate whether the type of linebreak that a system uses. Enumerated values are CR, LF, or CR/LF.
	QUALITY	Used to record a quality measure (as a string) for the output of the encoding process (OCR quality, transcription quality, etc.).
	role	Used to indicate the role of an agent. Enumerated values are OCR, TRANSCRIBER, MARKUP, and EDITOR.
	version	Used to record the version number (as a string) for a given piece of software, a markup language, or a schema version.

b) Digital images

There are several technical metadata standards for digital images, the most used being NISO Z39.87[13] and MIX, Adobe and XMP[14], EXif[15] and IPTC/XMP. For still digital images it is recommended to use NISO Data Dictionary – Technical Metadata for Digital Still Images or EXif Schema.

The technical metadata for digital images may include for example:

- File format
- File resolution (pixels per inch or other unit measurements)
- Dimensions (image dimension or size in inches or centimeters)
- Bit-depth (8-bit, 16-bit, 24-bit, etc.)
- Color mode (RGB, CMYK or grayscale)
- Scanner or Digital brand of manufacturer, name and model number
- Software used to manipulate or compress the image (including software number and version)

A detailed list of metadata elements, as well as their description, corresponding to EXif Schema is given in Table 17 from Appendix B. EXif Schema for digital still images at page 90.

c) Digital audio files

Long-term access to digitized audio files may depend on the quality of technical metadata captured during digitization. For digital audio files the technical metadata may include, for example:

- file format, including whether the format is lossless or lossy
- physical format for media-dependent materials
- sample rate
- track format for magnetic tape recordings
- number of channels and sound channel map
- sampling frequency and bits per sample for digital recordings
- audio data encoding for digital materials
- software used to manipulate or compress the audio file, including the software name and version
- brand, name, and model number of the recording equipment used

Table 7 presents several standards that are available for technical digital audio files.

[13] http://www.niso.org/apps/group_public/download.php/17936/z39-87-2006_2017.pdf

[14] http://www.adobe.com/products/xmp/standards.html

[15] https://www.w3.org/2003/12/exif/

Table 7. Metadata standards for digital audio

Standard	Description
AES57-2011	This standard provides a vocabulary to be used in describing structural and administrative metadata for digital and analog audio formats for the purpose of enabling audio preservation activities on those objects. The characteristics of the audio objects captured by this standard may be of use to audio communities beyond the audio preservation community.
AudioMD	AudioMD contains technical metadata that describe a digital audio archival object. Developed by the Library of Congress.
EAD Encoded Archival Description	EAD is a non-proprietary de facto standard for the encoding of finding aids for use in a networked (online) environment. Finding aids are inventories, indexes, or guides that are created by archival and manuscript repositories to provide information about specific collections. While the finding aids may vary somewhat in style, their common purpose is to provide detailed description of the content, history, and intellectual organization of collections of archival materials. EAD allows the standardization of collection information in finding aids within and across repositories.
EBUCore	The EBU Core Metadata Set is proposed to aggregate information within the EBU (European Broadcast Union) community. The Dublin Core used in the EBU Core Metadata Set will allow EBU Members to contribute to Europeana (the European Digital Library Project in collaboration with National Libraries and Museums), and EUScreen (a European project to provide access to broadcasters audiovisual resource with a participation of several EBU members and in liaison with Europeana).
PBCore	Version 2.0 of PBCore (the Public Broadcasting Metadata Dictionary) has been developed by a cross-organizational team of public radio and television producers and managers, archivists and information scientists to provide—for television, radio and Web activities—a standard way of describing and using descriptive data, allowing content to be more easily retrieved and shared among colleagues, software systems, institutions, community and production partners, private citizens, and educators. It can also be used as a guide for the onset of an archival or asset management process at an individual station or institution.
WAVE file embedded metadata	Embedded metadata is a part of the content file and cannot become separated from it. It can provide a brief statement of information provided in fuller form through external metadata standards. The Broadcast WAVE file provides a place in the file header for extended information about file creation, content, etc. Federal Agencies Audio-Visual Working Group, draft guideline on Broadcast WAVE Metadata represents a minimal set of recommended embedded metadata for historical and cultural heritage digital audio reformatting.
ISO/IEC 15938-3, MPEG-7 Audio	MPEG-7 Audio, standardizes the description tools for describing audio content. Most Audio description tools are based on audio features that let us measure similarity in sounds (such as music and speech). Therefore, MPEG-7 Audio descriptors and description schemes can be used to search and filter audio content based on several audio features like spectrum, harmony, timbre, and melody.

Currently the most used standard for technical audio metadata is AES57-2011[16] and was released by Audio Engineering Society (AES) in 2011 and updated in 2017. This standard covers audio object structures for preservation and restoration, both for analog and digital files. A detailed review of the standard is provided in the paper of Otto, J. *A Sound Strategy for Preservation: Adapting Audio Engineering Society Technical Metadata for Use in Multimedia Repositories*[17].

d) Digital video files

There are several technical metadata standards for digital video and multimedia files, the most common being presented in Table 8.

[16] http://www.aes.org/publications/standards/search.cfm?docID=84

[17] https://rucore.libraries.rutgers.edu/rutgers-lib/45764/PDF/1/play/

Table 8. Metadata standards for digital video files

Standard	Description
ISO/IEC 15398 - MPEG-7	This standard allows for extremely detailed recording of technical metadata as well as rights metadata and descriptive metadata for the intellectual content of the materials. Is the standard for describing multimedia content that provides the richest multimedia content description tools for applications ranging from content management, organization, navigation, and automated processing.
PBCore	Version 2.0 of PBCore (the Public Broadcasting Metadata Dictionary) has been developed by a cross-organizational team of public radio and television producers and managers, archivists and information scientists to provide—for television, radio and Web activities—a standard way of describing and using descriptive data, allowing content to be more easily retrieved and shared among colleagues, software systems, institutions, community and production partners, private citizens, and educators. It can also be used as a guide for the onset of an archival or asset management process at an individual station or institution.
reVTMD	This is an emerging standard designed by NARA to describe the process history of digitizing audiovisual materials. It allows for detailed technical description of every piece of hardware and software used in the digitization chain, as well as the role that each tool played. However, it has not been widely documented or adopted as of yet; its main use is as an extension in other schemas, as it is really the only existing schema that allows for technically detailed process history metadata.
videoMD	VideoMD was designed by the Library of Congress to provide technical metadata about video objects. It allows for a great deal of granularity and is a good choice for institutions that simply wish to describe technical information at the object level, rather than creating complex relationships between audiovisual objects. Much of the information described by videoMD can be automatically extracted from digital objects using programs like MediaInfo and FFprobe, then mapped into a METS file using the videoMD data structure.
UTVideo	UTVideo was designed by University of Texas to provide comprehensive metadata that will facilitate the discovery and preservation of fragile, born-digital video and is based on a data architecture which can be used in a METS-based or METS-compliant environment.

Two of the most commonly used standards for digital video file technical metadata are videoMD[18] and MPEG-7[19]. The videoMD schema allows for detailing properties such as:

- physical format for media dependent materials (e.g., stock brand, base and binder for magnetic tapes, disc surface type for disc media)
- dimensions for media dependent materials (e.g., diameter, gauge, height, thickness, width, length)

[18] https://www.loc.gov/standards/amdvmd/

[19] https://mpeg.chiariglione.org/standards/mpeg-7

- generation for analog media
- frame description (e.g., pixel or line counts horizontal and vertical, frame rate)
- data rate (may be expressed in terms of maximum, minimum, nominal, mode)
- format in terms of broadcast standards (e.g., NTSC, PAL, SECAM)
- formatting description for digital materials (e.g., name of creating application, commercial name of format, profile, version)
- bits per sample and sampling description in terms of chrominance and luminance

An example of videoMD technical metadata for a digital video file, extracted with MediaInfo software[20], is illustrated in Table 9.

Table 9. MediaInfo output for a digital video file (videoMD standard)

Type	Element	Value
General	Complete name	C:\Danubius-RI EN.mp4
	Format	MPEG-4
	Format profile	Base Media / Version 2
	Codec ID	mp42 (isom/mp42)
	File size	172 MiB
	Duration	14 min 1 s
	Overall bit rate mode	Variable
	Overall bit rate	1 714 kb/s
	Encoded date	UTC 2016-08-29 08:07:29
	Tagged date	UTC 2016-08-29 08:07:29
	gsst	0
	gstd	841885
Video	ID	1
	Format	AVC
	Format/Info	Advanced Video Codec
	Format profile	Main@L3.1

[20] https://mediaarea.net/en/MediaInfo

	Format settings	CABAC / 3 Ref Frames
	Format settings, CABAC	Yes
	Format settings, RefFrames	3 frames
	Codec ID	avc1
	Codec ID/Info	Advanced Video Coding
	Duration	14 min 1 s
	Bit rate	1 585 kb/s
	Width	1 280 pixels
	Height	720 pixels
	Display aspect ratio	16:9
	Frame rate mode	Constant
	Frame rate	25.000 FPS
	Color space	YUV
	Chroma subsampling	4:2:0
	Bit depth	8 bits
	Scan type	Progressive
	Bits/(Pixel*Frame)	0.069
	Stream size	159 MiB (92%)
	Title	ISO Media file produced by Google Inc.
	Encoded date	UTC 2016-08-29 08:07:29
	Tagged date	UTC 2016-08-29 08:07:29
	ID	2
Audio	Format	AAC
	Format/Info	Advanced Audio Codec
	Format profile	LC
	Codec ID	mp4a-40-2
	Duration	14 min 1 s

Bit rate mode	Variable
Bit rate	126 kb/s
Channel(s)	2 channels
Channel positions	Front: L R
Sampling rate	44.1 kHz
Frame rate	43.066 FPS (1024 SPF)
Compression mode	Lossy
Stream size	12.6 MiB (7%)
Title	ISO Media file produced by Google Inc.
Encoded date	UTC 2016-08-29 08:07:29
Tagged date	UTC 2016-08-29 08:07:29

The MPEG-7 standard defines a large library of core description tools, and a set of system tools provides the means for deploying the description in specific storage and transport environments[21]. MPEG-7 addresses many different applications in many different environments, which means it needs to provide a flexible framework for describing multimedia data, including extensibility (using the Description Definition Language) and restrictibility (via the MPEG-7 Profiles under specification).

The description tools used to describe video and image content follow ISO/IEC 15938-3 (MPEG-7 Visual) standard. Since there are no visual description schemes, the visual descriptors are based on visual features that measure similarity in images and videos (e.g. color, texture, shape, object motion, camera motion, etc.) and that can be employed in searching and filtering operations. Table 10 provides examples of generic visual descriptors for MPEG-7 Visual standard.

Table 10. Generic visual descriptors for MPEG-7 Visual standard

Visual descriptors	Examples
Basic elements	Grid Layout, Time Series, 2D–3D Multiple View, Spatial 2D Coordinates, Temporal Interpolation
Color descriptors	Color Space, Color Quantization, Scalable Color, Dominant Color, Color Layout, Color Structure, and Group-of-Frames / Group-of-Pictures Color

[21] MPEG-7: Overview of MPEG-7 Description Tools, José M. Martínez, Copyright © 2002 IEEE. Reprinted from IEEE Computer Society, July-September 2002

Texture descriptors	Homogeneous Texture, NonHomogeneous Texture (Edge histogram), Texture Browsing
Shape descriptors	Region-Based, Contour-Based, 3D Shape
Motion descriptors	Motion Activity, Camera Motion, Parametric Motion, Motion Trajectory
Location descriptors	Region Locator and Spatio–Temporal Locator

4.2. Preservation

Preservation metadata establishes an informational frame of reference around a preserved digital object that remains attached to that object over time. In order maintain the ability to exploit the full value of a preserved digital object into the future requires preserving this frame of reference in the form of well-maintained preservation metadata.

One way of interpreting this frame of reference uses the *provenance* of the object/resource. The provenance of a resource, according to the World Wide Web Consortium (W3C) Provenance Incubator Group[22], is "a record that describes entities and processes involved in producing and delivering or otherwise influencing that resource". Provenance information includes descriptions of the actions that have been taken to preserve the resource over time. Such information describes aspects of the digital preservation process used to maintain the object; it would also record any consequences of this process that alter the content, or look, feel, and functionality of the object. Related to this would be information that serves to establish and validate the object's authenticity. Provenance metadata is a mechanism to provide data about those entities, and their relationships to the resource and to other entities. The W3C group developed a provenance data model used to identify the characteristics of the object/resource or of the entities that have influenced it and the relationships between resources and entities.

This data model is based on three core structures: *entity*, *agent*, and *activity* defined as a resource, an entity that has influenced the life cycle of that resource and an activity is the nature of that influence (see Figure 5).

[22] https://www.w3.org/2011/prov/wiki/Main_Page

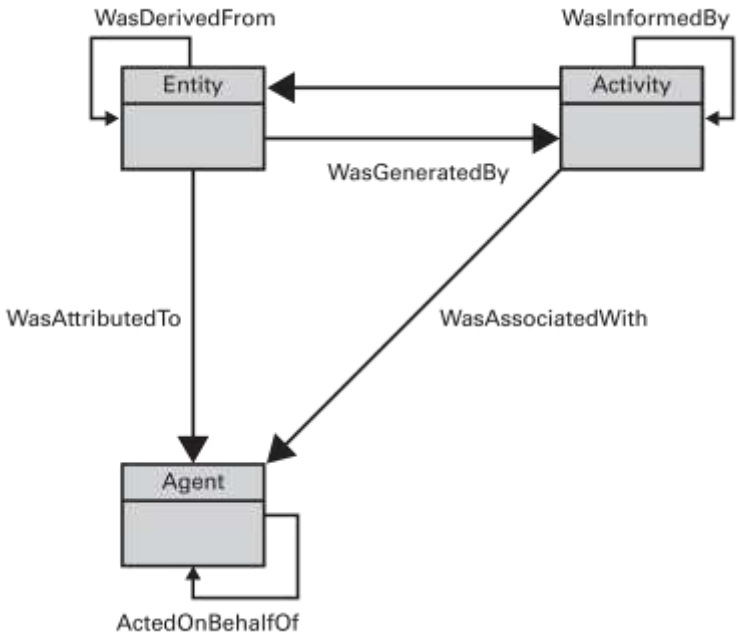

Figure 5. W3C provenance data model

One of the most developed metadata schema for supporting preservation is a standard from the Library of Congress: Preservation Metadata Implementation Strategies[23]. PREMIS was developed to be a core set of metadata elements for the preservation of digital objects. The PREMIS Data Dictionary defines preservation metadata as "the information a repository uses to support the digital preservation process". Some examples of preservation activities are given bellow:

- A resource must be stored securely so that nobody can modify it inadvertently (or maliciously). Checksum information stored as metadata can be used to tell if a stored file has changed between two points in time.
- Files must be stored on media that can be read by current computers. If the media are damaged or obsolete it can be difficult or impossible to recover the data. Metadata can support media management by recording the type and age of storage media and the dates that files were last refreshed.
- Over long periods of time even popular file formats can become obsolete, meaning no current applications can render them. Preservation managers must employ preservation strategies to ensure the resources remain usable.
- Preservation actions may entail changing original resources or changing how they are rendered. Metadata can help support authenticity by documenting the digital provenance of the resource -- its chain of custody and authorized change history

The PREMIS data model defines four entities of importance to the preservation process: *objects* (digital resources, which may be abstract intellectual entities), *agents* (people or organizations that may influence the object), *events* (time-stamped actions performed by agents on the object), and *rights statements* (permissions such as intellectual property rights).

[23] https://www.loc.gov/standards/premis/index.html

Each of these entities contains a set of *semantic units* which in other metadata schemas would be called elements.

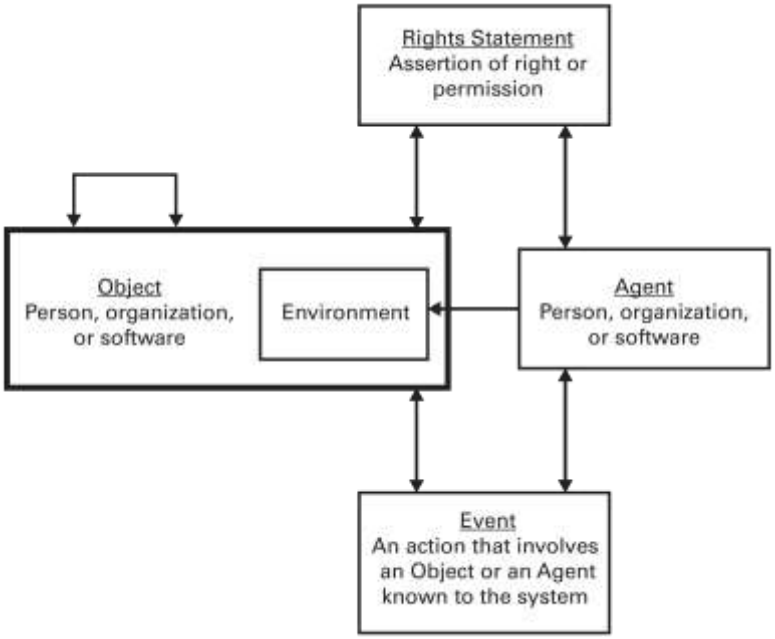

Figure 6. PREMIS data model

PREMIS specifies a large number of semantic units for these four entities. A full and detailed description of each semantic unit can be found in Appendix A. Premis Data Dictionary semantic units at https://www.loc.gov/standards/premis/v3/premis-3-0-datadictionary-only.pdf.

4.3. Rights

Rights metadata describes copyright and the terms of use of the resource described. The importance of rights metadata has increased due to easy transfer and reuse of digital content. Dublin Core provides a minimal number of terms for capturing data about rights (e.g. *license*, *rightsHolder* or *accessRights* – for more see Table 5), all of which are quite broad, with no recommended best practices for how to select or construct values.

The Getty Research Institute[24] proposed a *rights metadata dictionary* whose goal was to simplify the rights metadata understanding by clarifying which work was being described in a record and allowing users to better understand some of the ambiguities or unknowns about the rights information provided. They stated that in order to obtain usable, shareable, repurposable rights metadata one needs to include the core information (Table 11):

Table 11. Core elements for rights metadata (the Getty Research Institute rights metadata dictionary)[25]

Element	Valid data values
Title	The data values for this element should be copied (preferably in an automated manner) from the title element from the descriptive

[24] https://www.getty.edu

[25] http://www.getty.edu/research/publications/electronic_publications/intrometadata/rights.pdf

Element	Valid data values
	metadata record for the work or item. This element, which is repeatable, can contain translated titles, brief titles, display titles, etc., in addition to the title that is inscribed on the item or object, if one exists.
Creator	The name of the creator of the original object or work, taken from a published controlled vocabulary or local authority file whenever possible. The life dates in the case of individual creators, including the death date if applicable. Dates should be expressed according to a standard format, (e.g., ISO 8601).
Creation dates	The date(s) of the creation of the work. Dates should be expressed according to a standard format (e.g., ISO 8601).
Creator nationality	The nationality or culture of the creator of the work, if known
Copyright status	Valid values for this element should be selected from a controlled list. • *Copyright owned by the institution* that holds the original object/work or item • *Copyright owned by a third party* - include a subelement for the name of the third party, taken from a published controlled vocabulary whenever possible • *Public domain* • *Orphan work* • *Not yet researched*
Publication status	Valid values for this element should be selected from a controlled list. For example: • *Published* - include a subelement with the date of publication, if known, in a standard format (e.g., ISO 8601). Note that date of creation and date of publication are not necessarily identical. • *Unpublished* (in which case, the creator dates and/or date of creation are extremely important) • *Unknown*, after research and due diligence • *Not yet researched*
Date of rights metadata research	This should be a repeating element, since metadata research is often necessarily an incremental process to which more than one individual contributes. The individual's name or initials should be provided by the information system and associated with the relevant dates of research. Dates should be expressed according to a standard format (e.g., ISO 8601).

A more complete rights metadata schema is the METSRights Rights Declaration Schema[26], or RightsDeclarationMD. This schema was developed to be an extension to the Metadata Encoding and Transmission Standard (METS)[27]. METS was developed to fill the need for a standard data structure for describing complex digital library objects. METS is an XML Schema for creating XML document instances that express the structure of digital library objects, the associated descriptive and administrative metadata, and the names and locations of the files that comprise the digital object. Expressed using the XML schema language, METS provides a document format for encoding the metadata necessary for management of digital library objects within a repository and for exchange between repositories.

RightsDeclarationMD has three top-level elements: *RightsDeclaration* (the rights associated with a resource), *RightsHolder* (an individual or organization), and *Context* (a description of what rights holders have what rights, and under what circumstances). Each of these top-level elements has several attributes, for example, an attribute of *RightsDeclaration* is *RightsCategory*, which may be populated with the values from a small controlled vocabulary that includes copyrighted, licensed, public domain, etc.; an attribute of the *Context* element is *Permissions*, which also has a small controlled vocabulary associated with it, including such values as discover, display, copy, modify, and delete.

5. Metadata standards

Standards are developed because of a need to connect systems together. It is a search for compatibility between the systems. There are many standards available, and many more that are in development. Choosing the best ones for your system can be a challenge, because there is a huge width of alternatives. It is sometimes difficult to select standards that are compatible with systems and technologies that already exist. The standards continue to evolve whilst the systems might be deployed and running for years. We all need standards to ensure that systems and software are interoperable.

5.1. Standards by discipline

There are many different metadata schemes being developed in a variety of user environments and disciplines. The following table presents several of the most used metadata standards according to their discipline:

[26] http://www.loc.gov/standards/mets/news080503.html

[27] http://www.loc.gov/standards/mets/mets-home.html

Table 12. Standards of metadata[28]

Domain	Standards	Description
Arts and Humanities	***DDI (Data Documentation Initiative)***	A widely used, international standard for describing data from the social, behavioral, and economic sciences. Two versions of the standard are currently maintained in parallel: DDI Codebook (or DDI version 2) is the simpler of the two, and intended for documenting simple survey data for exchange or archiving. Version 2.5 was released in January 2014.DDI Lifecycle (or DDI version 3) is richer and may be used to document datasets at each stage of their lifecycle from conceptualization through to publication and reuse. It is modular and extensible. Version 3.2 was published in March 2014.Both versions are XML-based and defined using XML Schemas. They were developed and are maintained by the DDI Alliance.
	MIDAS-Heritage	A British cultural heritage standard for recording information on buildings, archaeological sites, shipwrecks, parks and gardens, battlefields, areas of interest and artefacts. Sponsored by the Forum on Information Standards in Heritage, MIDAS Version 1.1 was released in October 2012.
	OAI-ORE (Open Archives Initiative Object Reuse and Exchange)	The goal of these standards is to expose the rich content in aggregations of Web resources to applications that support authoring, deposit, exchange, visualization, reuse, and preservation. The standards support the changing nature of scholarship and scholarly communication, and the need for cyberinfrastructure to support that scholarship, with the intent to develop standards that generalize across all web-based information including the increasing popular social networks of “Web 2.0”.
Engineering	***CIF (Crystallographic Information Framework)***	A well-established standard file structure for the archiving and distribution of crystallographic information, CIF is in regular use for reporting crystal structure determinations to Acta Crystallographica and other journals. Sponsored by the International Union of Crystallography, the current standard dates from 1997. As of July 2011, a new version of the CIF standard is under consideration.
	CSMD (Core Scientific Metadata Model)	A study-data oriented model, primarily in support of the ICAT data managment infrastructure software. The CSMD is designed to support data collected within a large-scale facility’s scientific workflow; however, the model is also designed to be generic across scientific disciplines. Sponsored by the Science and Technologies Facilities Council, the latest full specification available is v 4.0, from 2013.
	ISA-Tab	The Investigation/Study/Assay (ISA) tab-delimited (TAB) format is a general-purpose framework with which to collect and communicate complex metadata (i.e. sample

[28] https://www.getty.edu/publications/intrometadata/

Domain	Standards	Description
		characteristics, technologies used, type of measurements made) from 'omics-based' experiments employing a combination of technologies. Created by core developers from the University of Oxford, ISA-TAB v1.0 was released in November 2008.
	MIBBI (Minimum Information for Biological and Biomedical Investigations)	A common portal to a group of nearly 40 checklists of Minimum Information for various biological disciplines. The MIBBI Foundry is developing a cross-analysis of these guidelines to create an intercompatible, extensible community of standards. The concept was realized initially through the joint efforts of the Proteomics Standards Initiative, the Genomic Standards Consortium and the MGED RSBI Working Groups. The latest project to register with MIBBI is the MIABie guidelines for reporting biofilm research, as of January 2012.
	NeXus	NeXus is an international standard for the storage and exchange of neutron, x-ray, and muon experiment data. The structure of NeXus files is extremely flexible, allowing the storage of both simple data sets, such as a single data array and its axes, and highly complex data and their associated metadata, such as measurements on a multi-component instrument or numerical simulations. NeXus is built on top of the container format HDF5, and adds domain-specific rules for organizing data within HDF5 files in addition to a dictionary of well-defined domain-specific field names.
Life Sciences	***ABCD (Access to Biological Collection Data)***	The Access to Biological Collections Data (ABCD) Schema is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data). The ABCD Schema attempts to be comprehensive and highly structured, supporting data from a wide variety of databases. It is compatible with several existing data standards. Parallel structures exist so that either (or both) atomised data and free-text can be accommodated. Sponsored by Biodiversity Information Standards TDWG - the Taxonomic Databases Working Group, the current specification was last modified in 2007.
	Darwin Core	A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries. Sponsored by Biodiversity Information Standards (TWDG), the current standard was last modified in October 2009.
	EML (Ecological Metadata Language)	Ecological Metadata Language (EML) is a metadata specification particularly developed for the ecology discipline. It is based on prior work done by the Ecological Society of America and associated efforts (Michener et al., 1997, Ecological Applications). Sponsored by

Domain	Standards	Description
		ecoinformatics.org, EML Version 2.1.1 was released in 2011.
	Genome Metadata	Genome metadata on PATRIC consists of 61 different metadata fields, called attributes, which are organized into the following seven broad categories: Organism Info, Isolate Info, Host Info, Sequence Info, Phenotype Info, Project Info, and Others.
	ISA-Tab	The Investigation/Study/Assay (ISA) tab-delimited (TAB) format is a general-purpose framework with which to collect and communicate complex metadata (i.e. sample characteristics, technologies used, type of measurements made) from 'omics-based' experiments employing a combination of technologies. Created by core developers from the University of Oxford, ISA-TAB v1.0 was released in November 2008.
	MIBBI (Minimum Information for Biological and Biomedical Investigations)	A common portal to a group of nearly 40 checklists of Minimum Information for various biological disciplines. The MIBBI Foundry is developing a cross-analysis of these guidelines to create an intercompatible, extensible community of standards. The concept was realized initially through the joint efforts of the Proteomics Standards Initiative, the Genomic Standards Consortium and the MGED RSBI Working Groups. The latest project to register with MIBBI is the MIABie guidelines for reporting biofilm research, as of January 2012.
	NeXus	NeXus is an international standard for the storage and exchange of neutron, x-ray, and muon experiment data. The structure of NeXus files is extremely flexible, allowing the storage of both simple data sets, such as a single data array and its axes, and highly complex data and their associated metadata, such as measurements on a multi-component instrument or numerical simulations. NeXus is built on top of the container format HDF5, and adds domain-specific rules for organizing data within HDF5 files in addition to a dictionary of well-defined domain-specific field names.
	Observ-OM	Observ-OM is founded on four basic concepts to represent any kind of observation: Targets, Features, Protocols (and their Applications), and Values. It is intended to lower the barrier for future data sharing and facilitate integrated search across panels and species. All models, formats, documentation, and software are available for free and open source (LGPLv3) at http://www.observ-om.org.
	OME-XML (Open Microscopy Environment XML)	OME-XML is a vendor-neutral file format for biological image data, with an emphasis on metadata supporting light microscopy. It can be used as a data file format in its own right, or as a way of encoding metadata within a TIFF or BigTIFF file (for which purpose there is the OME-TIFF specification). The standard is maintained by the Open Microscopy Environment Consortium, and was last updated in June 2012.

Domain	Standards	Description
	PDBx/mmCIF (Protein Data Bank Exchange Dictionary and the Macromolecular Crystallographic Information Framework)	Protein Data Bank archive (PDB) is the single worldwide archival repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies, managed by the Worldwide PDB (wwPDB). The PDB Exchange Dictionary (PDBx) is used by the wwPDB to define data content for deposition, annotation and archiving of PDB entries. PDBx incorporates the community standard metadata representation, the Macromolecular Crystallographic Information Framework (mmCIF), orginally developed under the auspices of the International Union of Crystallography (IUCr). PDBx has been extended by the wwPDB to include descriptions of other experimental methods that produce 3D macromolecular structure models such as Nuclear Magnetic Resonance Spectroscopy, 3D Electron Microscopy and Tomography.
	Protocol Data Element Definitions	A draft set of data elements required by the National Institues of Health (U.S.) for the submission of trial information to the CLincalTrials.gov registry and results database.
	Repository-Developed Metadata Schemas	Some repositories have decided that current standards do not fit their metadata needs, and so have created their own requirements.
	UKEOF	A metadata standard for describing environmental monitoring activities, programmes, networks and facilities published by the UK Environmental Observation Framework (UKEOF).
Physical Sciences & Mathematics	***AgMES (Agricultural Metadata Element Set)***	A semantic standard developed by the Food and Agriculture Organization (FAO) of the United Nations, AgMES enables description, resource discovery, interoperability and data exchange of different types of information resources in all areas relevant to food production, nutrition and rural development. Sponsored by the UN AIMS - Agricultural Information Managment Standards, the current standard was issued in November 2010.
	AVM (Astronomy Visualization Metadata)	The AVM scheme supports the cross-searching of collections of print-ready and screen-ready astronomical imagery rendered from telescopic observations (also known as ‘pretty pictures’). The scheme is compatible with the Adobe XMP specification, so the metadata can be embedded within common image formats such as JPEG, TIFF and PNG.Such images can combine data acquired at different wavebands and from different observatories. While the primary intent is to cover data-derived astronomical images, there are broader uses as well. Specifically, the most general subset of this schema is also appropriate for describing artwork and illustrations of astronomical subject matter.AVM is a proposed recommendation of the International Virtual Observatory Alliance and was last updated in 2011.

Domain	Standards	Description
	CF (Climate and Forecast) Metadata Conventions	The CF standard was originally framed as a standard for data written in netCDF format, with model-generated climate forecast data particularly in mind. However, it is equally applicable to observational datasets, and can be used to describe other formats. It is a standard for "use metadata" that aims both to distinguish quantities (such as physical description, units, and prior processing) and to locate the data in space–time. Sponsored by the NetCDF Climate and Forecast Metadata Convention, the current version dates from December 2011.
	CIF (Crystallographic Information Framework)	A well-established standard file structure for the archiving and distribution of crystallographic information, CIF is in regular use for reporting crystal structure determinations to Acta Crystallographica and other journals. Sponsored by the International Union of Crystallography, the current standard dates from 1997. As of July 2011, a new version of the CIF standard is under consideration.
	CIM (Common Information Model)	The Common Information Model (CIM) describes climate data, the models and software from which they derive, the geographic grids used to calculate and project them, and the experimental processes (typically simulations) that produced them. The CIM was originally developed by the EU-funded Metafor Project. It is now maintained and developed by Earth Science Documentation (ES-DOC). The latest release dates from 2012.
	CSMD (Core Scientific Metadata Model)	A study-data oriented model, primarily in support of the ICAT data managment infrastructure software. The CSMD is designed to support data collected within a large-scale facility's scientific workflow; however, the model is also designed to be generic across scientific disciplines. Sponsored by the Science and Technologies Facilities Council, the latest full specification available is v 4.0, from 2013.
	DIF (Directory Interchange Format)	An early metadata initiative from the Earth sciences community, intended for the description of scientific data sets. It inlcudes elements focusing on instruments that capture data, temporal and spatial characteristics of the data, and projects with which the dataset is associated. It is defined as a W3C XML Schema.Sponsored by the Global Change Master Directory; the DIF Writer's Guide Version 6 is from November 2010.
	FGDC/CSDGM (Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata)	A widely-used, but no longer current standard defining the information content for a set of digital geospatial data required by the US Federal Government.CSDGM was sponsored by the US Federal Geographic Data Committee. However, in September 2010 the FGDC endorsed ISO 19115 and began encouraging federal agencies to transition to ISO metadata.
	FITS (Flexible Image Transport System)	FITS is an image data file format for encoding astronomical data. The WCS (World Coordinate System) conventions map elements in data arrays to standard

Domain	Standards	Description
		physical coordinates in the sky. FITS has provisions for image metadata encoded in an ASCII header at the beginning of files.
	International Virtual Observatory Alliance Technical Specifications	The technical specifications defined by the IVOA (International Virtual Observatory Alliance) enable interoperability between and the integration of astronomical archives across the world into an international virtual observatory. They include several data models that act as metadata schemas for particular data types: for example, photometry data, simulation data, space-time coordinates, spectral lines data, spectral data, observational data, and the physical parameter space of astronomical datasets.These data models are under active development by the IVOA Data Modelling Working Group.Additional recommendations have been made for metadata concepts and terms necessary for the discovery and the use of astronomical data collections and services.
	ISO 19115	An internationally-adopted schema for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. Sponsored by the International Standards Organisation, the first edition of ISO 19115 was published in 2003. It has since been split into parts: ISO 19115-1:2014 contains the fundamentals of the standard; ISO 19115-2:2009 contains extensions for imagery and gridded data; and ISO/TS 19115-3:2016 provides an XML schema implementation for the fundamental concepts compatible with ISO/TS 19138:2007 (Geographic Metadata XML, or GMD).
	NeXus	NeXus is an international standard for the storage and exchange of neutron, x-ray, and muon experiment data. The structure of NeXus files is extremely flexible, allowing the storage of both simple data sets, such as a single data array and its axes, and highly complex data and their associated metadata, such as measurements on a multi-component instrument or numerical simulations. NeXus is built on top of the container format HDF5, and adds domain-specific rules for organizing data within HDF5 files in addition to a dictionary of well-defined domain-specific field names.
	Observations and Measurements	This encoding is an essential dependency for the OGC Sensor Observation Service (SOS) Interface Standard. More specifically, this standard defines XML schemas for observations, and for features involved in sampling when making observations. These provide document models for the exchange of information describing observation acts and their results, both within and between different scientific and technical communities.
	PDBx/mmCIF (Protein Data Bank Exchange Dictionary	Protein Data Bank archive (PDB) is the single worldwide archival repository of information about the 3D structures

Domain	Standards	Description
	and the Macromolecular Crystallographic Information Framework)	of proteins, nucleic acids, and complex assemblies, managed by the Worldwide PDB (wwPDB). The PDB Exchange Dictionary (PDBx) is used by the wwPDB to define data content for deposition, annotation and archiving of PDB entries. PDBx incorporates the community standard metadata representation, the Macromolecular Crystallographic Information Framework (mmCIF), orginally developed under the auspices of the International Union of Crystallography (IUCr). PDBx has been extended by the wwPDB to include descriptions of other experimental methods that produce 3D macromolecular structure models such as Nuclear Magnetic Resonance Spectroscopy, 3D Electron Microscopy and Tomography.
	Repository-Developed Metadata Schemas	Some repositories have decided that current standards do not fit their metadata needs, and so have created their own requirements.
	SDAC (Standard for Documentation of Astronomical Catalogues)	The Standard for Documentation of Astronomical Catalogues is a set of conventions for archiving astronomical data. As well as path, filename and data format conventions, it also specifies how to construct a plain text description file for documenting the data files. It was developed as an alternative to FITS that would be more suited to archives, permit human inspection, and allow manipulation via standard Unix command-line tools. SDAC was developed by CDS (Centre de Données astronomiques de Strasbourg). Version 2.0 is the most recent; it was released in February 2000.
	SPASE Data Model	An information model for describing the elements of the heliophysics data environment, and a set of resource types which can be used to describe data along with its scientific context, source, provenance, content and location. It is designed to support a federated data system where data may reside at different locations and may be seperated from the metadata which describes it. The preferred expression form is XML.The Space Physics Archive Search and Extract (SPASE) effort is implemented by the SPASE Consortium which is composed of representatives of the international Heliophysics data community. The Current Release of the data model (2.2.2) was updated in October 2012.
	UKEOF	A metadata standard for describing environmental monitoring activities, programmes, networks and facilities published by the UK Environmental Observation Framework (UKEOF).
Social and Behavioral Sciences	***DDI (Data Documentation Initiative)***	A widely used, international standard for describing data from the social, behavioral, and economic sciences. Two versions of the standard are currently maintained in parallel: DDI Codebook (or DDI version 2) is the simpler of the two, and intended for documenting simple survey data for exchange or archiving. Version 2.5 was released in January 2014.DDI Lifecycle (or DDI version 3) is richer

Domain	Standards	Description
		and may be used to document datasets at each stage of their lifecycle from conceptualization through to publication and reuse. It is modular and extensible. Version 3.2 was published in March 2014.Both versions are XML-based and defined using XML Schemas. They were developed and are maintained by the DDI Alliance.
	MIDAS-Heritage	A British cultural heritage standard for recording information on buildings, archaeological sites, shipwrecks, parks and gardens, battlefields, areas of interest and artefacts. Sponsored by the Forum on Information Standards in Heritage, MIDAS Version 1.1 was released in October 2012.
	OAI-ORE (Open Archives Initiative Object Reuse and Exchange)	The goal of these standards is to expose the rich content in aggregations of Web resources to applications that support authoring, deposit, exchange, visualization, reuse, and preservation. The standards support the changing nature of scholarship and scholarly communication, and the need for cyberinfrastructure to support that scholarship, with the intent to develop standards that generalize across all web-based information including the increasing popular social networks of “Web 2.0”.
	QuDEx (Qualitative Data Exchange Format)	The QuDEx standard/schema is a software-neutral format for qualitative data that preserves annotations of, and relationships between, data and other related objects. It can be viewed as the optimal baseline data exchange model for the archiving and interchange of data and metadata.
	SDMX (Statistical Data and Metadata Exchange)	A set of common technical and statistical standards and guidelines to be used for the efficient exchange and sharing of statistical data and metadata. Sponsoring institutions include BIS, ECB, EUROSTAT, IMF, OECD, UN, and the World Bank. Technical Specification 2.1 was amended in May 2012.
General Research Data	***CERIF (Common European Research Information Format)***	The Common European Research Information Format is the standard that the EU recommends to its member states for recording information about research activity. Since version 1.6 it has included specific support for recording metadata for datasets.
	Data Package	The Data Package specification is a generic wrapper format for exchanging data. Although it supports arbitrary metadata, the format defines required, recommended, and optional fields for both the package as a whole and the resources contained within it.A separate but linked specification provides a way to describe the columns of a data table; descriptions of this form can be included directly in the Data Package metadata.
	DataCite Metadata Schema	A set of mandatory metadata that must be registered with the DataCite Metadata Store when minting a DOI persistent identifier for a dataset. The domain-agnostic properties were chosen for their ability to aid in accurate

Domain	Standards	Description
		and consistent identification of data for citation and retrieval purposes. Sponsored by the DataCite consortium, version 3.0 was recently released in 2013.
	DCAT (Data Catalog Vocabulary)	By using DCAT to describe datasets in data catalogs, publishers increase discoverability and enable applications easily to consume metadata from multiple catalogs. It further enables decentralized publishing of catalogs and facilitates federated dataset search across sites. Aggregated DCAT metadata can serve as a manifest file to facilitate digital preservation.
	Dublin Core	A basic, domain-agnostic standard which can be easily understood and implemented, and as such is one of the best known and most widely used metadata standards. Sponsored by the Dublin Core Metadata Initiative, Dublin Core was published as ISO Standard 15836 in February 2009.
	OAI-ORE (Open Archives Initiative Object Reuse and Exchange)	The goal of these standards is to expose the rich content in aggregations of Web resources to applications that support authoring, deposit, exchange, visualization, reuse, and preservation. The standards support the changing nature of scholarship and scholarly communication, and the need for cyberinfrastructure to support that scholarship, with the intent to develop standards that generalize across all web-based information including the increasing popular social networks of “Web 2.0”.
	Observations and Measurements	This encoding is an essential dependency for the OGC Sensor Observation Service (SOS) Interface Standard. More specifically, this standard defines XML schemas for observations, and for features involved in sampling when making observations. These provide document models for the exchange of information describing observation acts and their results, both within and between different scientific and technical communities.
	PREMIS	The PREMIS (Preservation Metadata: Implementation Strategies) Data Dictionary defines a set of metadata that most repositories of digital objects would need to record and use in order to preserve those objects over the long term. It has its roots in the Open Archival Information System Reference Model but has been strongly influenced by the practical experience of such repositories. While the Data Dictionary can be used with other standards to influence the creation of local application profiles, an XML Schema is provided to allow the metadata to be serialized independently. PREMIS was initially developed by the Preservation Metadata: Implementation Strategies Working Group, convened by OCLC and RLG, and is currently maintained by the PREMIS Maintenance Activity, lead by the Library of Congress.
	PROV	Provenance is information about entities, activities, and people involved in producing a piece of data or thing,

Domain	Standards	Description
		which can be used to form assessments about its quality, reliability or trustworthiness. The PROV Family of Documents defines a model, corresponding serializations and other supporting definitions to enable the inter-operable interchange of provenance information in heterogeneous environments such as the Web.
	RDF Data Cube Vocabulary	The standard provides a means to publish multi-dimensional data, such as statistics, on the web in such a way that it can be linked to related data sets and concepts using the W3C RDF (Resource Description Framework) standard. The model underpinning the Data Cube vocabulary is compatible with the cube model that underlies SDMX (Statistical Data and Metadata eXchange), an ISO standard for exchanging and sharing statistical data and metadata among organizations.
	Repository-Developed Metadata Schemas	Some repositories have decided that current standards do not fit their metadata needs, and so have created their own requirements.

5.2. Standards by subject areas

The DANUBIUS research infrastructure will occupy a central position within the strategic research landscape of river-sea systems. The services and applications that DANUBIUS-RI will provide, cover both public and private sectors and are extendable to several domains as depicted in Figure 7.

The European Commission, the United States National Science Foundation and National Institute of Standards and Technology, and the Australian Government's Department of Innovation launched in 2013 an organization called Research Data Alliance (RDA[29]) focused on the development of infrastructure and community activities that reduce barriers to data sharing and exchange, and the acceleration of data driven innovation worldwide.

Table 13 presents metadata standards for different subject areas within the described domains according to RD-Alliance organisation.

[29] https://www.rd-alliance.org/about-rda

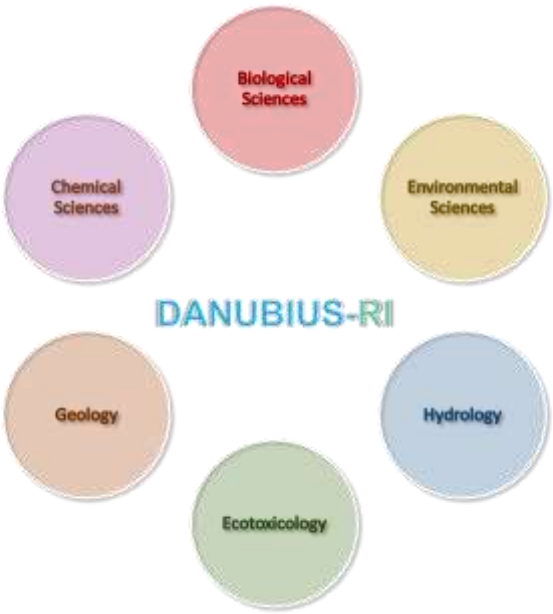

Figure 7. Intersection domains of DANUBIUS-RI

Table 13. Metadata standards by subject areas according to RD-Alliance project

Subject	Standard	Content	Website	Specification
Agricultural Science	*AgMES (Agricultural Metadata Element Set)*	A semantic standard developed by the Food and Agriculture Organization (FAO) of the United Nations, AgMES enables description, resource discovery, interoperability and data exchange of different types of information resources in all areas relevant to food production, nutrition and rural development. Sponsored by the UN AIMS - Agricultural Information Managment Standards, the current standard was issued in November 2010.	http://aims.fao.org/standards/agmes	http://aims.fao.org/es/standards/agmes/namespace-specification
	DIF (Directory Interchange Format)	An early metadata initiative from the Earth sciences community, intended for the description of scientific data sets. It inlcudes elements focusing on instruments that capture data, temporal and spatial characteristics of the data, and projects with which the dataset is associated. It is defined as a W3C XML Schema.Sponsored by the Global Change Master Directory, the DIF Writer's Guide Version 6 is from November 2010.	http://gcmd.nasa.gov/add/difguide/	http://gcmd.nasa.gov/add/difguide/WRITEADIF.pdf
Animal pathology	*OME-XML (Open Microscopy Environment XML)*	OME-XML is a vendor-neutral file format for biological image data, with an emphasis on metadata supporting light microscopy. It can be used as a data file format in its own right, or as a way of encoding metadata within a TIFF or BigTIFF file (for which purpose there is the OME-TIFF specification). The standard is maintained by the Open Microscopy Environment Consortium, and was last updated in June 2012.	http://www.openmicroscopy.org/site/support/file-formats	http://www.openmicroscopy.org/Schemas/
	OME-XML (Open Microscopy Environment XML)	OME-XML is a vendor-neutral file format for biological image data, with an emphasis on metadata supporting light microscopy. It can be used as a data file format in its own right, or as a way of encoding metadata within a TIFF or BigTIFF file (for which purpose there is the OME-TIFF specification). The standard is maintained by the Open Microscopy Environment Consortium, and was last updated in June 2012.	http://www.openmicroscopy.org/site/support/file-formats	http://www.openmicroscopy.org/Schemas/

Subject	Standard	Content	Website	Specification
Biochemistry	*CSMD (Core Scientific Metadata Model)*	A study-data oriented model, primarily in support of the ICAT data managment infrastructure software. The CSMD is designed to support data collected within a large-scale facility's scientific workflow; however the model is also designed to be generic across scientific disciplines.Sponsored by the Science and Technologies Facilities Council, the latest full specification available is v 4.0, from 2013.	http://icatproject-contrib.github.io/CSMD/	http://purl.org/net/CSMD/
	DIF (Directory Interchange Format)	An early metadata initiative from the Earth sciences community, intended for the description of scientific data sets. It inlcudes elements focusing on instruments that capture data, temporal and spatial characteristics of the data, and projects with which the dataset is associated. It is defined as a W3C XML Schema.Sponsored by the Global Change Master Directory, the DIF Writer's Guide Version 6 is from November 2010.	http://gcmd.nasa.gov/add/difguide/	http://gcmd.nasa.gov/add/difguide/WRITEADIF.pdf
	FGDC/CSDGM (Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata)	A widely-used, but no longer current standard defining the information content for a set of digital geospatial data required by the US Federal Government.CSDGM was sponsored by the US Federal Geographic Data Committee. However, in September 2010 the FGDC endorsed ISO 19115 and began encouraging federal agencies to transition to ISO metadata.	http://www.fgdc.gov/metadata/geospatial-metadata-standards/	http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf
	ISA-Tab	The Investigation/Study/Assay (ISA) tab-delimited (TAB) format is a general purpose framework with which to collect and communicate complex metadata (i.e. sample characteristics, technologies used, type of measurements made) from 'omics-based' experiments employing a combination of technologies.Created by core developers from the University of Oxford, ISA-TAB v1.0 was released in November 2008.	http://isatab.sourceforge.net/format.html	http://isatab.sourceforge.net/docs/ISA-TAB_release-candidate-1_v1.0_24nov08.pdf

Subject	Standard	Content	Website	Specification
Biodiversity	*MIBBI (Minimum Information for Biological and Biomedical Investigations)*	A common portal to a group of nearly 40 checklists of Minimum Information for various biological disciplines. The MIBBI Foundry is developing a cross-analysis of these guidelines to create an intercompatible, extensible community of standards.The concept was realized initially through the joint efforts of the Proteomics Standards Initiative, the Genomic Standards Consortium and the MGED RSBI Working Groups. The latest project to register with MIBBI is the MIABie guidelines for reporting biofilm research, as of January 2012.	http://mibbi.sourceforge.net/portal.shtml	http://mibbi.sourceforge.net/foundry.shtml
	ABCD (Access to Biological Collection Data)	The Access to Biological Collections Data (ABCD) Schema is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data). The ABCD Schema attempts to be comprehensive and highly structured, supporting data from a wide variety of databases. It is compatible with several existing data standards. Parallel structures exist so that either (or both) atomised data and free-text can be accommodated.Sponsored by Biodiversity Information Standards TDWG - the Taxonomic Databases Working Group, the current specification was last modified in 2007.	http://wiki.tdwg.org/ABCD	http://www.tdwg.org/standards/115/
	Darwin Core	A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries.Sponsored by Biodiversity Information Standards (TWDG), the current standard was last modified in October 2009.	http://rs.tdwg.org/dwc/index.htm	http://www.tdwg.org/standards/450
	EML (Ecological Metadata Language)	Ecological Metadata Language (EML) is a metadata specification particularly developed for the ecology discipline. It is based on prior work done by the Ecological Society of America and associated efforts (Michener et al., 1997, Ecological Applications).Sponsored by ecoinformatics.org, EML Version 2.1.1 was released in 2011.	http://knb.ecoinformatics.org/software/eml/	http://knb.ecoinformatics.org/software/eml/eml-2.1.1/index.html

Subject	Standard	Content	Website	Specification
	UKEOF	A metadata standard for describing environmental monitoring activities, programmes, networks and facilities published by the UK Environmental Observation Framework (UKEOF).	http://www.ukeof.org.uk/	http://schema.ukeof.org.uk/
Biogeography	*Darwin Core*	A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries.Sponsored by Biodiversity Information Standards (TWDG), the current standard was last modified in October 2009.	http://rs.tdwg.org/dwc/index.htm	http://www.tdwg.org/standards/450
	UKEOF	A metadata standard for describing environmental monitoring activities, programmes, networks and facilities published by the UK Environmental Observation Framework (UKEOF).	http://www.ukeof.org.uk/	http://schema.ukeof.org.uk/
Bioinformatics	*MIBBI (Minimum Information for Biological and Biomedical Investigations)*	A common portal to a group of nearly 40 checklists of Minimum Information for various biological disciplines. The MIBBI Foundry is developing a cross-analysis of these guidelines to create an intercompatible, extensible community of standards.The concept was realized initially through the joint efforts of the Proteomics Standards Initiative, the Genomic Standards Consortium and the MGED RSBI Working Groups. The latest project to register with MIBBI is the MIABie guidelines for reporting biofilm research, as of January 2012.	http://mibbi.sourceforge.net/portal.shtml	http://mibbi.sourceforge.net/foundry.shtml
	PDBx/mmCIF (Protein Data Bank Exchange Dictionary and the Macromolecular Crystallographic	Protein Data Bank archive (PDB) is the single worldwide archival repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies, managed by the Worldwide PDB (wwPDB). The PDB Exchange Dictionary (PDBx) is used by the wwPDB to define data content for deposition, annotation and archiving of PDB entries. PDBx incorporates the community standard metadata representation, the Macromolecular Crystallographic Information Framework (mmCIF), orginally developed under the auspices of the International Union of Crystallography (IUCr). PDBx	http://mmcif.wwpdb.org/	http://mmcif.wwpdb.org/dictionaries/mmcif_pdbx_v40.dic/Index/

Subject	Standard	Content	Website	Specification
	Information Framework)	has been extended by the wwPDB to include descriptions of other experimental methods that produce 3D macromolecular structure models such as Nuclear Magnetic Resonance Spectroscopy, 3D Electron Microscopy and Tomography.		
Biology	*Observ-OM*	Observ-OM is founded on four basic concepts to represent any kind of observation: Targets, Features, Protocols (and their Applications), and Values. It is intended to lower the barrier for future data sharing and facilitate integrated search across panels and species. All models, formats, documentation, and software are available for free and open source (LGPLv3) at http://www.observ-om.org.	http://www.molgenis.org/wiki/ObservStart	http://www.molgenis.org/wiki/ObservStart
	PDBx/mmCIF (Protein Data Bank Exchange Dictionary and the Macromolecular Crystallographic Information Framework)	Protein Data Bank archive (PDB) is the single worldwide archival repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies, managed by the Worldwide PDB (wwPDB). The PDB Exchange Dictionary (PDBx) is used by the wwPDB to define data content for deposition, annotation and archiving of PDB entries. PDBx incorporates the community standard metadata representation, the Macromolecular Crystallographic Information Framework (mmCIF), orginally developed under the auspices of the International Union of Crystallography (IUCr). PDBx has been extended by the wwPDB to include descriptions of other experimental methods that produce 3D macromolecular structure models such as Nuclear Magnetic Resonance Spectroscopy, 3D Electron Microscopy and Tomography.	http://mmcif.wwpdb.org/	http://mmcif.wwpdb.org/dictionaries/mmcif_pdbx_v40.dic/Index/
Biomaterials	*ISA-Tab*	The Investigation/Study/Assay (ISA) tab-delimited (TAB) format is a general purpose framework with which to collect and communicate complex metadata (i.e. sample characteristics, technologies used, type of measurements made) from 'omics-based' experiments employing a combination of technologies.Created by core developers from the University of Oxford, ISA-TAB v1.0 was released in November 2008.	http://isatab.sourceforge.net/format.html	http://isatab.sourceforge.net/docs/ISA-TAB_release-candidate-1_v1.0_24nov08.pdf

Subject	Standard	Content	Website	Specification
Botany	*ABCD (Access to Biological Collection Data)*	The Access to Biological Collections Data (ABCD) Schema is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data). The ABCD Schema attempts to be comprehensive and highly structured, supporting data from a wide variety of databases. It is compatible with several existing data standards. Parallel structures exist so that either (or both) atomised data and free-text can be accommodated.Sponsored by Biodiversity Information Standards TDWG - the Taxonomic Databases Working Group, the current specification was last modified in 2007.	http://wiki.tdwg.org/ABCD	http://www.tdwg.org/standards/115/
	Darwin Core	A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries.Sponsored by Biodiversity Information Standards (TWDG), the current standard was last modified in October 2009.	http://rs.tdwg.org/dwc/index.htm	http://www.tdwg.org/standards/450
	EML (Ecological Metadata Language)	Ecological Metadata Language (EML) is a metadata specification particularly developed for the ecology discipline. It is based on prior work done by the Ecological Society of America and associated efforts (Michener et al., 1997, Ecological Applications).Sponsored by ecoinformatics.org, EML Version 2.1.1 was released in 2011.	http://knb.ecoinformatics.org/software/eml/	http://knb.ecoinformatics.org/software/eml/eml-2.1.1/index.html
Cartography	*Darwin Core*	A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries.Sponsored by Biodiversity Information Standards (TWDG), the current standard was last modified in October 2009.	http://rs.tdwg.org/dwc/index.htm	http://www.tdwg.org/standards/450

Subject	Standard	Content	Website	Specification
	FGDC/CSDGM (Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata)	A widely-used, but no longer current standard defining the information content for a set of digital geospatial data required by the US Federal Government.CSDGM was sponsored by the US Federal Geographic Data Committee. However, in September 2010 the FGDC endorsed ISO 19115 and began encouraging federal agencies to transition to ISO metadata.	http://www.fgdc.gov/metadata/geospatial-metadata-standards/	http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf
	ISO 19115	An internationally-adopted schema for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data.Sponsored by the International Standards Organisation, the first edition of ISO 19115 was published in 2003. It has since been split into parts: ISO 19115-1:2014 contains the fundamentals of the standard; ISO 19115-2:2009 contains extensions for imagery and gridded data; and ISO/TS 19115-3:2016 provides an XML schema implementation for the fundamental concepts compatible with ISO/TS 19138:2007 (Geographic Metadata XML, or GMD).	http://www.iso.org/iso/catalogue_detail.htm?csnumber=53798	http://standards.iso.org/iso/19115/-3/
Cell Biology	*ISA-Tab*	The Investigation/Study/Assay (ISA) tab-delimited (TAB) format is a general purpose framework with which to collect and communicate complex metadata (i.e. sample characteristics, technologies used, type of measurements made) from 'omics-based' experiments employing a combination of technologies.Created by core developers from the University of Oxford, ISA-TAB v1.0 was released in November 2008.	http://isatab.sourceforge.net/format.html	http://isatab.sourceforge.net/docs/ISA-TAB_release-candidate-1_v1.0_24nov08.pdf
	MIBBI (Minimum Information for Biological and	A common portal to a group of nearly 40 checklists of Minimum Information for various biological disciplines. The MIBBI Foundry is developing a cross-analysis of these guidelines to create an	http://mibbi.sourceforge.net/portal.shtml	http://mibbi.sourceforge.net/foundry.shtml

Subject	Standard	Content	Website	Specification
	Biomedical Investigations)	intercompatible, extensible community of standards.The concept was realized initially through the joint efforts of the Proteomics Standards Initiative, the Genomic Standards Consortium and the MGED RSBI Working Groups. The latest project to register with MIBBI is the MIABie guidelines for reporting biofilm research, as of January 2012.		
	OME-XML (Open Microscopy Environment XML)	OME-XML is a vendor-neutral file format for biological image data, with an emphasis on metadata supporting light microscopy. It can be used as a data file format in its own right, or as a way of encoding metadata within a TIFF or BigTIFF file (for which purpose there is the OME-TIFF specification).The standard is maintained by the Open Microscopy Environment Consortium, and was last updated in June 2012.	http://www.openmicroscopy.org/site/support/file-formats	http://www.openmicroscopy.org/Schemas/
Chemistry	*CIF (Crystallographic Information Framework)*	A well-established standard file structure for the archiving and distribution of crystallographic information, CIF is in regular use for reporting crystal structure determinations to Acta Crystallographica and other journals.Sponsored by the International Union of Crystallography, the current standard dates from 1997. As of July 2011, a new version of the CIF standard is under consideration.	http://www.iucr.org/resources/cif	http://www.iucr.org/resources/cif/spec
	CSMD (Core Scientific Metadata Model)	A study-data oriented model, primarily in support of the ICAT data managment infrastructure software. The CSMD is designed to support data collected within a large-scale facility's scientific workflow; however the model is also designed to be generic across scientific disciplines.Sponsored by the Science and Technologies Facilities Council, the latest full specification available is v 4.0, from 2013.	http://icatproject-contrib.github.io/CSMD/	http://purl.org/net/CSMD/
	NeXus	NeXus is an international standard for the storage and exchange of neutron, x-ray, and muon experiment data. The structure of NeXus files is extremely flexible, allowing the storage of both simple data sets, such as a single data array and its axes, and highly complex data and their associated metadata, such as measurements on a	http://www.nexusformat.org/	http://download.nexusformat.org/doc/html/ref_doc.html

Subject	Standard	Content	Website	Specification
		multi-component instrument or numerical simulations. NeXus is built on top of the container format HDF5, and adds domain-specific rules for organizing data within HDF5 files in addition to a dictionary of well-defined domain-specific field names.		
	Observations and Measurements	This encoding is an essential dependency for the OGC Sensor Observation Service (SOS) Interface Standard. More specifically, this standard defines XML schemas for observations, and for features involved in sampling when making observations. These provide document models for the exchange of information describing observation acts and their results, both within and between different scientific and technical communities.	http://www.opengeospatial.org/standards/om	http://portal.opengeospatial.org/files/?artifact_id=41579
	PDBx/mmCIF (Protein Data Bank Exchange Dictionary and the Macromolecular Crystallographic Information Framework)	Protein Data Bank archive (PDB) is the single worldwide archival repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies, managed by the Worldwide PDB (wwPDB). The PDB Exchange Dictionary (PDBx) is used by the wwPDB to define data content for deposition, annotation and archiving of PDB entries. PDBx incorporates the community standard metadata representation, the Macromolecular Crystallographic Information Framework (mmCIF), orginally developed under the auspices of the International Union of Crystallography (IUCr). PDBx has been extended by the wwPDB to include descriptions of other experimental methods that produce 3D macromolecular structure models such as Nuclear Magnetic Resonance Spectroscopy, 3D Electron Microscopy and Tomography.	http://mmcif.wwpdb.org/	http://mmcif.wwpdb.org/dictionaries/mmcif_pdbx_v40.dic/Index/
Climatology	*CF (Climate and Forecast) Metadata Conventions*	The CF standard was originally framed as a standard for data written in netCDF format, with model-generated climate forecast data particularly in mind. However, it is equally applicable to observational datasets, and can be used to describe other formats. It is a standard for “use metadata” that aims both to distinguish quantities (such as physical description, units, and prior processing) and to locate the	http://cfconventions.org/	http://cfconventions.org/documents.html

Subject	Standard	Content	Website	Specification
		data in space–time.Sponsored by the NetCDF Climate and Forecast Metadata Convention, the current version dates from December 2011.		
	CIM (Common Information Model)	The Common Information Model (CIM) describes climate data, the models and software from which they derive, the geographic grids used to calculate and project them, and the experimental processes (typically simulations) that produced them.The CIM was originally developed by the EU-funded Metafor Project. It is now maintained and developed by Earth Science Documentation (ES-DOC). The latest release dates from 2012.	https://earthsystemcog.org/projects/es-doc-models/cim	https://earthsystemcog.org/projects/es-doc-models/cim_versions
	DIF (Directory Interchange Format)	An early metadata initiative from the Earth sciences community, intended for the description of scientific data sets. It inlcudes elements focusing on instruments that capture data, temporal and spatial characteristics of the data, and projects with which the dataset is associated. It is defined as a W3C XML Schema.Sponsored by the Global Change Master Directory, the DIF Writer's Guide Version 6 is from November 2010.	http://gcmd.nasa.gov/add/difguide/	http://gcmd.nasa.gov/add/difguide/WRITEADIF.pdf
	FGDC/CSDGM (Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata)	A widely-used, but no longer current standard defining the information content for a set of digital geospatial data required by the US Federal Government.CSDGM was sponsored by the US Federal Geographic Data Committee. However, in September 2010 the FGDC endorsed ISO 19115 and began encouraging federal agencies to transition to ISO metadata.	http://www.fgdc.gov/metadata/geospatial-metadata-standards/	http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf
	ISO 19115	An internationally-adopted schema for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema,	http://www.iso.org/iso/catalogue_detail.htm?csnumber=53798	http://standards.iso.org/iso/19115/-3/

Subject	Standard	Content	Website	Specification
		spatial reference, and distribution of digital geographic data.Sponsored by the International Standards Organisation, the first edition of ISO 19115 was published in 2003. It has since been split into parts: ISO 19115-1:2014 contains the fundamentals of the standard; ISO 19115-2:2009 contains extensions for imagery and gridded data; and ISO/TS 19115-3:2016 provides an XML schema implementation for the fundamental concepts compatible with ISO/TS 19138:2007 (Geographic Metadata XML, or GMD).		
	WMO CDM4	Climate Metadata of WMO's World Climate Data and Monitoring Programme (WCDMP). WCDMP is a sub-program of the World Climate Programme (WCP), provides an international coordination of the WMO Climate System Monitoring. Priorities of WCDMP include the provision of guidance to the WMO Members in the area of climate data and climate assessment and monitoring; and the publication and dissemination of WMO authoritative reports on the status of the global climate. GUIDELINES ON CLIMATE METADATA AND HOMOGENIZATION by Enric Aguilar1, Inge Auer, et al.	http://www.wmo.int/pages/prog/wcp/wcdmp/index_en.php	http://www.wmo.int/pages/prog/wcp/wcdmp/documents/WCDMP-53.pdf
	UKEOF	A metadata standard for describing environmental monitoring activities, programmes, networks and facilities published by the UK Environmental Observation Framework (UKEOF).	http://www.ukeof.org.uk/	http://schema.ukeof.org.uk/
Ecology	*Darwin Core*	A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries.Sponsored by Biodiversity Information Standards (TWDG), the current standard was last modified in October 2009.	http://rs.tdwg.org/dwc/index.htm	http://www.tdwg.org/standards/450

Subject	Standard	Content	Website	Specification
	DIF (Directory Interchange Format)	An early metadata initiative from the Earth sciences community, intended for the description of scientific data sets. It inlcudes elements focusing on instruments that capture data, temporal and spatial characteristics of the data, and projects with which the dataset is associated. It is defined as a W3C XML Schema.Sponsored by the Global Change Master Directory, the DIF Writer's Guide Version 6 is from November 2010.	http://gcmd.nasa.gov/add/difguide/	http://gcmd.nasa.gov/add/difguide/WRITEADIF.pdf
	EML (Ecological Metadata Language)	Ecological Metadata Language (EML) is a metadata specification particularly developed for the ecology discipline. It is based on prior work done by the Ecological Society of America and associated efforts (Michener et al., 1997, Ecological Applications).Sponsored by ecoinformatics.org, EML Version 2.1.1 was released in 2011.	http://knb.ecoinformatics.org/software/eml/	http://knb.ecoinformatics.org/software/eml/eml-2.1.1/index.html
	UKEOF	A metadata standard for describing environmental monitoring activities, programmes, networks and facilities published by the UK Environmental Observation Framework (UKEOF).	http://www.ukeof.org.uk/	http://schema.ukeof.org.uk/
Environmental Science	*MIBBI (Minimum Information for Biological and Biomedical Investigations)*	A common portal to a group of nearly 40 checklists of Minimum Information for various biological disciplines. The MIBBI Foundry is developing a cross-analysis of these guidelines to create an intercompatible, extensible community of standards. The concept was realized initially through the joint efforts of the Proteomics Standards Initiative, the Genomic Standards Consortium and the MGED RSBI Working Groups. The latest project to register with MIBBI is the MIABie guidelines for reporting biofilm research, as of January 2012.	http://mibbi.sourceforge.net/portal.shtml	http://mibbi.sourceforge.net/foundry.shtml

Subject	Standard	Content	Website	Specification
Fish Farming	*AgMES (Agricultural Metadata Element Set)*	A semantic standard developed by the Food and Agriculture Organization (FAO) of the United Nations, AgMES enables description, resource discovery, interoperability and data exchange of different types of information resources in all areas relevant to food production, nutrition and rural development. Sponsored by the UN AIMS - Agricultural Information Managment Standards, the current standard was issued in November 2010.	http://aims.fao.org/standards/agmes	http://aims.fao.org/es/standards/agmes/namespace-specification
Genetics	*Genome Metadata*	Genome metadata on PATRIC consists of 61 different metadata fields, called attributes, which are organized into the following seven broad categories: Organism Info, Isolate Info, Host Info, Sequence Info, Phenotype Info, Project Info, and Others.	http://enews.patricbrc.org/faqs/genome-metadata-faqs/	http://enews.patricbrc.org/faqs/genome-metadata-faqs/
	ISA-Tab	The Investigation/Study/Assay (ISA) tab-delimited (TAB) format is a general purpose framework with which to collect and communicate complex metadata (i.e. sample characteristics, technologies used, type of measurements made) from 'omics-based' experiments employing a combination of technologies. Created by core developers from the University of Oxford, ISA-TAB v1.0 was released in November 2008.	http://isatab.sourceforge.net/format.html	http://isatab.sourceforge.net/docs/ISA-TAB_release-candidate-1_v1.0_24nov08.pdf
	MIBBI (Minimum Information for Biological and Biomedical Investigations)	A common portal to a group of nearly 40 checklists of Minimum Information for various biological disciplines. The MIBBI Foundry is developing a cross-analysis of these guidelines to create an intercompatible, extensible community of standards. The concept was realized initially through the joint efforts of the Proteomics Standards Initiative, the Genomic Standards Consortium and the MGED RSBI Working Groups. The latest project to register with MIBBI is the MIABie guidelines for reporting biofilm research, as of January 2012.	http://mibbi.sourceforge.net/portal.shtml	http://mibbi.sourceforge.net/foundry.shtml

Subject	Standard	Content	Website	Specification
Genomics	*ABCD (Access to Biological Collection Data)*	The Access to Biological Collections Data (ABCD) Schema is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data). The ABCD Schema attempts to be comprehensive and highly structured, supporting data from a wide variety of databases. It is compatible with several existing data standards. Parallel structures exist so that either (or both) atomised data and free-text can be accommodated. Sponsored by Biodiversity Information Standards TDWG - the Taxonomic Databases Working Group, the current specification was last modified in 2007.	http://wiki.tdwg.org/ABCD	http://www.tdwg.org/standards/115/
	Genome Metadata	Genome metadata on PATRIC consists of 61 different metadata fields, called attributes, which are organized into the following seven broad categories: Organism Info, Isolate Info, Host Info, Sequence Info, Phenotype Info, Project Info, and Others.	http://enews.patricbrc.org/faqs/genome-metadata-faqs/	http://enews.patricbrc.org/faqs/genome-metadata-faqs/
	ISA-Tab	The Investigation/Study/Assay (ISA) tab-delimited (TAB) format is a general-purpose framework with which to collect and communicate complex metadata (i.e. sample characteristics, technologies used, type of measurements made) from 'omics-based' experiments employing a combination of technologies. Created by core developers from the University of Oxford, ISA-TAB v1.0 was released in November 2008.	http://isatab.sourceforge.net/format.html	http://isatab.sourceforge.net/docs/ISA-TAB_release-candidate-1_v1.0_24nov08.pdf
	MIBBI (Minimum Information for Biological and Biomedical Investigations)	A common portal to a group of nearly 40 checklists of Minimum Information for various biological disciplines. The MIBBI Foundry is developing a cross-analysis of these guidelines to create an intercompatible, extensible community of standards. The concept was realized initially through the joint efforts of the Proteomics Standards Initiative, the Genomic Standards Consortium and the MGED RSBI Working Groups. The latest project to register with MIBBI is the MIABie guidelines for reporting biofilm research, as of January 2012.	http://mibbi.sourceforge.net/portal.shtml	http://mibbi.sourceforge.net/foundry.shtml

Subject	Standard	Content	Website	Specification
Geography	*FGDC/CSDGM (Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata)*	A widely-used, but no longer current standard defining the information content for a set of digital geospatial data required by the US Federal Government.CSDGM was sponsored by the US Federal Geographic Data Committee. However, in September 2010 the FGDC endorsed ISO 19115 and began encouraging federal agencies to transition to ISO metadata.	http://www.fgdc.gov/metadata/geospatial-metadata-standards/	http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf
	ISO 19115	An internationally-adopted schema for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. Sponsored by the International Standards Organisation, the first edition of ISO 19115 was published in 2003. It has since been split into parts: ISO 19115-1:2014 contains the fundamentals of the standard; ISO 19115-2:2009 contains extensions for imagery and gridded data; and ISO/TS 19115-3:2016 provides an XML schema implementation for the fundamental concepts compatible with ISO/TS 19138:2007 (Geographic Metadata XML, or GMD).	http://www.iso.org/iso/catalogue_detail.htm?csnumber=53798	http://standards.iso.org/iso/19115/-3/
Geology	*ABCD (Access to Biological Collection Data)*	The Access to Biological Collections Data (ABCD) Schema is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data). The ABCD Schema attempts to be comprehensive and highly structured, supporting data from a wide variety of databases. It is compatible with several existing data standards. Parallel structures exist so that either (or both) atomised data and free-text can be accommodated. Sponsored by Biodiversity Information Standards TDWG - the Taxonomic Databases Working Group, the current specification was last modified in 2007.	http://wiki.tdwg.org/ABCD	http://www.tdwg.org/standards/115/

Subject	Standard	Content	Website	Specification
	ISO 19115	An internationally-adopted schema for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. Sponsored by the International Standards Organisation, the first edition of ISO 19115 was published in 2003. It has since been split into parts: ISO 19115-1:2014 contains the fundamentals of the standard; ISO 19115-2:2009 contains extensions for imagery and gridded data; and ISO/TS 19115-3:2016 provides an XML schema implementation for the fundamental concepts compatible with ISO/TS 19138:2007 (Geographic Metadata XML, or GMD).	http://www.iso.org/iso/catalogue_detail.htm?csnumber=53798	http://standards.iso.org/iso/19115/-3/
	Observations and Measurements	This encoding is an essential dependency for the OGC Sensor Observation Service (SOS) Interface Standard. More specifically, this standard defines XML schemas for observations, and for features involved in sampling when making observations. These provide document models for the exchange of information describing observation acts and their results, both within and between different scientific and technical communities.	http://www.opengeospatial.org/standards/om	http://portal.opengeospatial.org/files/?artifact_id=41579
	UKEOF	A metadata standard for describing environmental monitoring activities, programmes, networks and facilities published by the UK Environmental Observation Framework (UKEOF).	http://www.ukeof.org.uk/	http://schema.ukeof.org.uk/
Geoscience	*ABCD (Access to Biological Collection Data)*	The Access to Biological Collections Data (ABCD) Schema is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data). The ABCD Schema attempts to be comprehensive and highly structured, supporting data from a wide variety of databases. It is compatible with several existing data standards. Parallel structures exist so that either (or both) atomised data and free-text can be accommodated. Sponsored by Biodiversity Information Standards	http://wiki.tdwg.org/ABCD	http://www.tdwg.org/standards/115/

Subject	Standard	Content	Website	Specification
		TDWG - the Taxonomic Databases Working Group, the current specification was last modified in 2007.		
	Darwin Core	A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries. Sponsored by Biodiversity Information Standards (TWDG), the current standard was last modified in October 2009.	http://rs.tdwg.org/dwc/index.htm	http://www.tdwg.org/standards/450
	ISO 19115	An internationally-adopted schema for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. Sponsored by the International Standards Organisation, the first edition of ISO 19115 was published in 2003. It has since been split into parts: ISO 19115-1:2014 contains the fundamentals of the standard; ISO 19115-2:2009 contains extensions for imagery and gridded data; and ISO/TS 19115-3:2016 provides an XML schema implementation for the fundamental concepts compatible with ISO/TS 19138:2007 (Geographic Metadata XML, or GMD).	http://www.iso.org/iso/catalogue_detail.htm?csnumber=53798	http://standards.iso.org/iso/19115/-3/
Hydrogeology	*ISO 19115*	An internationally-adopted schema for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. Sponsored by the International Standards Organisation, the first edition of ISO 19115 was published in 2003. It has since been split into parts: ISO 19115-1:2014 contains the fundamentals of the standard; ISO 19115-2:2009 contains extensions for imagery and gridded data; and ISO/TS 19115-3:2016 provides an XML schema implementation for the fundamental concepts compatible with ISO/TS 19138:2007 (Geographic Metadata XML, or GMD).	http://www.iso.org/iso/catalogue_detail.htm?csnumber=53798	http://standards.iso.org/iso/19115/-3/

Subject	Standard	Content	Website	Specification
Hydrography	*DIF (Directory Interchange Format)*	An early metadata initiative from the Earth sciences community, intended for the description of scientific data sets. It inlcudes elements focusing on instruments that capture data, temporal and spatial characteristics of the data, and projects with which the dataset is associated. It is defined as a W3C XML Schema.Sponsored by the Global Change Master Directory, the DIF Writer's Guide Version 6 is from November 2010.	http://gcmd.nasa.gov/add/difguide/	http://gcmd.nasa.gov/add/difguide/WRITEADIF.pdf
	FGDC/CSDGM (Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata)	A widely-used, but no longer current standard defining the information content for a set of digital geospatial data required by the US Federal Government.CSDGM was sponsored by the US Federal Geographic Data Committee. However, in September 2010 the FGDC endorsed ISO 19115 and began encouraging federal agencies to transition to ISO metadata.	http://www.fgdc.gov/metadata/geospatial-metadata-standards/	http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf
	ISO 19115	An internationally-adopted schema for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. Sponsored by the International Standards Organisation, the first edition of ISO 19115 was published in 2003. It has since been split into parts: ISO 19115-1:2014 contains the fundamentals of the standard; ISO 19115-2:2009 contains extensions for imagery and gridded data; and ISO/TS 19115-3:2016 provides an XML schema implementation for the fundamental concepts compatible with ISO/TS 19138:2007 (Geographic Metadata XML, or GMD).	http://www.iso.org/iso/catalogue_detail.htm?csnumber=53798	http://standards.iso.org/iso/19115/-3/

Subject	Standard	Content	Website	Specification
Hydrology	*DIF (Directory Interchange Format)*	An early metadata initiative from the Earth sciences community, intended for the description of scientific data sets. It inlcudes elements focusing on instruments that capture data, temporal and spatial characteristics of the data, and projects with which the dataset is associated. It is defined as a W3C XML Schema.Sponsored by the Global Change Master Directory, the DIF Writer's Guide Version 6 is from November 2010.	http://gcmd.nasa.gov/add/difguide/	http://gcmd.nasa.gov/add/difguide/WRITEADIF.pdf
	ISO 19115	An internationally-adopted schema for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. Sponsored by the International Standards Organisation, the first edition of ISO 19115 was published in 2003. It has since been split into parts: ISO 19115-1:2014 contains the fundamentals of the standard; ISO 19115-2:2009 contains extensions for imagery and gridded data; and ISO/TS 19115-3:2016 provides an XML schema implementation for the fundamental concepts compatible with ISO/TS 19138:2007 (Geographic Metadata XML, or GMD).	http://www.iso.org/iso/catalogue_detail.htm?csnumber=53798	http://standards.iso.org/iso/19115/-3/
	UKEOF	A metadata standard for describing environmental monitoring activities, programmes, networks and facilities published by the UK Environmental Observation Framework (UKEOF).	http://www.ukeof.org.uk/	http://schema.ukeof.org.uk/
Marine Biology	*DIF (Directory Interchange Format)*	An early metadata initiative from the Earth sciences community, intended for the description of scientific data sets. It inlcudes elements focusing on instruments that capture data, temporal and spatial characteristics of the data, and projects with which the dataset is associated. It is defined as a W3C XML Schema.Sponsored by the Global Change Master Directory, the DIF Writer's Guide Version 6 is from November 2010.	http://gcmd.nasa.gov/add/difguide/	http://gcmd.nasa.gov/add/difguide/WRITEADIF.pdf

Subject	Standard	Content	Website	Specification
Marine Science	*FGDC/CSDGM (Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata)*	A widely-used, but no longer current standard defining the information content for a set of digital geospatial data required by the US Federal Government.CSDGM was sponsored by the US Federal Geographic Data Committee. However, in September 2010 the FGDC endorsed ISO 19115 and began encouraging federal agencies to transition to ISO metadata.	http://www.fgdc.gov/metadata/geospatial-metadata-standards/	http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf
	CF (Climate and Forecast) Metadata Conventions	The CF standard was originally framed as a standard for data written in netCDF format, with model-generated climate forecast data particularly in mind. However, it is equally applicable to observational datasets, and can be used to describe other formats. It is a standard for "use metadata" that aims both to distinguish quantities (such as physical description, units, and prior processing) and to locate the data in space–time. Sponsored by the NetCDF Climate and Forecast Metadata Convention, the current version dates from December 2011.	http://cfconventions.org/	http://cfconventions.org/documents.html
	Darwin Core	A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries. Sponsored by Biodiversity Information Standards (TWDG), the current standard was last modified in October 2009.	http://rs.tdwg.org/dwc/index.htm	http://www.tdwg.org/standards/450
	DIF (Directory Interchange Format)	An early metadata initiative from the Earth sciences community, intended for the description of scientific data sets. It inlcudes elements focusing on instruments that capture data, temporal and spatial characteristics of the data, and projects with which the dataset is associated. It is defined as a W3C XML Schema.Sponsored by the	http://gcmd.nasa.gov/add/difguide/	http://gcmd.nasa.gov/add/difguide/WRITEADIF.pdf

Subject	Standard	Content	Website	Specification
		Global Change Master Directory, the DIF Writer's Guide Version 6 is from November 2010.		
	FGDC/CSDGM (Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata)	A widely-used, but no longer current standard defining the information content for a set of digital geospatial data required by the US Federal Government.CSDGM was sponsored by the US Federal Geographic Data Committee. However, in September 2010 the FGDC endorsed ISO 19115 and began encouraging federal agencies to transition to ISO metadata.	http://www.fgdc.gov/metadata/geospatial-metadata-standards/	http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf
	ISO 19115	An internationally-adopted schema for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. Sponsored by the International Standards Organisation, the first edition of ISO 19115 was published in 2003. It has since been split into parts: ISO 19115-1:2014 contains the fundamentals of the standard; ISO 19115-2:2009 contains extensions for imagery and gridded data; and ISO/TS 19115-3:2016 provides an XML schema implementation for the fundamental concepts compatible with ISO/TS 19138:2007 (Geographic Metadata XML, or GMD).	http://www.iso.org/iso/catalogue_detail.htm?csnumber=53798	http://standards.iso.org/iso/19115/-3/
Marine Zoology	*Darwin Core*	A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries. Sponsored by Biodiversity Information Standards (TWDG), the current standard was last modified in October 2009.	http://rs.tdwg.org/dwc/index.htm	http://www.tdwg.org/standards/450

Subject	Standard	Content	Website	Specification
Maritime Geography	*Darwin Core*	A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries. Sponsored by Biodiversity Information Standards (TWDG), the current standard was last modified in October 2009.	http://rs.tdwg.org/dwc/index.htm	http://www.tdwg.org/standards/450
Metabolic biochemistry	*ISA-Tab*	The Investigation/Study/Assay (ISA) tab-delimited (TAB) format is a general-purpose framework with which to collect and communicate complex metadata (i.e. sample characteristics, technologies used, type of measurements made) from 'omics-based' experiments employing a combination of technologies. Created by core developers from the University of Oxford, ISA-TAB v1.0 was released in November 2008.	http://isatab.sourceforge.net/format.html	http://isatab.sourceforge.net/docs/ISA-TAB_release-candidate-1_v1.0_24nov08.pdf
	MIBBI (Minimum Information for Biological and Biomedical Investigations)	A common portal to a group of nearly 40 checklists of Minimum Information for various biological disciplines. The MIBBI Foundry is developing a cross-analysis of these guidelines to create an intercompatible, extensible community of standards. The concept was realized initially through the joint efforts of the Proteomics Standards Initiative, the Genomic Standards Consortium and the MGED RSBI Working Groups. The latest project to register with MIBBI is the MIABie guidelines for reporting biofilm research, as of January 2012.	http://mibbi.sourceforge.net/portal.shtml	http://mibbi.sourceforge.net/foundry.shtml
Molecular biology	*ISA-Tab*	The Investigation/Study/Assay (ISA) tab-delimited (TAB) format is a general-purpose framework with which to collect and communicate complex metadata (i.e. sample characteristics, technologies used, type of measurements made) from 'omics-based' experiments employing a combination of technologies. Created by core developers from the University of Oxford, ISA-TAB v1.0 was released in November 2008.	http://isatab.sourceforge.net/format.html	http://isatab.sourceforge.net/docs/ISA-TAB_release-candidate-1_v1.0_24nov08.pdf

Subject	Standard	Content	Website	Specification
Multi-disciplinary	*PDBx/mmCIF (Protein Data Bank Exchange Dictionary and the Macromolecular Crystallographic Information Framework)*	Protein Data Bank archive (PDB) is the single worldwide archival repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies, managed by the Worldwide PDB (wwPDB). The PDB Exchange Dictionary (PDBx) is used by the wwPDB to define data content for deposition, annotation and archiving of PDB entries. PDBx incorporates the community standard metadata representation, the Macromolecular Crystallographic Information Framework (mmCIF), orginally developed under the auspices of the International Union of Crystallography (IUCr). PDBx has been extended by the wwPDB to include descriptions of other experimental methods that produce 3D macromolecular structure models such as Nuclear Magnetic Resonance Spectroscopy, 3D Electron Microscopy and Tomography.	http://mmcif.wwpdb.org/	http://mmcif.wwpdb.org/dictionaries/mmcif_pdbx_v40.dic/Index/
	CERIF (Common European Research Information Format)	The Common European Research Information Format is the standard that the EU recommends to its member states for recording information about research activity. Since version 1.6 it has included specific support for recording metadata for datasets.	http://www.eurocris.org/cerif/main-features-cerif	http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.6/documentation/MInfo.html
	Data Package	The Data Package specification is a generic wrapper format for exchanging data. Although it supports arbitrary metadata, the format defines required, recommended, and optional fields for both the package as a whole and the resources contained within it.A separate but linked specification provides a way to describe the columns of a data table; descriptions of this form can be included directly in the Data Package metadata.	http://frictionlessdata.io/data-packages/	http://specs.frictionlessdata.io/data-packages/
	DataCite Metadata Schema	A set of mandatory metadata that must be registered with the DataCite Metadata Store when minting a DOI persistent identifier for a dataset. The domain-agnostic properties were chosen for their ability to aid in accurate and consistent identification of data for	http://schema.datacite.org	http://schema.datacite.org/meta/kernel-3/index.html

Subject	Standard	Content	Website	Specification
		citation and retrieval purposes. Sponsored by the DataCite consortium, version 3.0 was recently released in 2013.		
	Dublin Core	A basic, domain-agnostic standard which can be easily understood and implemented, and as such is one of the best known and most widely used metadata standards. Sponsored by the Dublin Core Metadata Initiative, Dublin Core was published as ISO Standard 15836 in February 2009.	http://dublincore.org	http://dublincore.org/specifications/
	OAI-ORE (Open Archives Initiative Object Reuse and Exchange)	The goal of these standards is to expose the rich content in aggregations of Web resources to applications that support authoring, deposit, exchange, visualization, reuse, and preservation. The standards support the changing nature of scholarship and scholarly communication, and the need for cyberinfrastructure to support that scholarship, with the intent to develop standards that generalize across all web-based information including the increasing popular social networks of “Web 2.0”.	http://www.openarchives.org/ore/toc	http://www.openarchives.org/ore/
Oceanography	*CF (Climate and Forecast) Metadata Conventions*	The CF standard was originally framed as a standard for data written in netCDF format, with model-generated climate forecast data particularly in mind. However, it is equally applicable to observational datasets, and can be used to describe other formats. It is a standard for “use metadata” that aims both to distinguish quantities (such as physical description, units, and prior processing) and to locate the data in space–time. Sponsored by the NetCDF Climate and Forecast Metadata Convention, the current version dates from December 2011.	http://cfconventions.org/	http://cfconventions.org/documents.html
	DIF (Directory Interchange Format)	An early metadata initiative from the Earth sciences community, intended for the description of scientific data sets. It inlcudes elements focusing on instruments that capture data, temporal and spatial characteristics of the data, and projects with which the dataset is associated. It is defined as a W3C XML Schema.Sponsored by the	http://gcmd.nasa.gov/add/difguide/	http://gcmd.nasa.gov/add/difguide/WRITEADIF.pdf

Subject	Standard	Content	Website	Specification
		Global Change Master Directory, the DIF Writer's Guide Version 6 is from November 2010.		
	FGDC/CSDGM (Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata)	A widely-used, but no longer current standard defining the information content for a set of digital geospatial data required by the US Federal Government.CSDGM was sponsored by the US Federal Geographic Data Committee. However, in September 2010 the FGDC endorsed ISO 19115 and began encouraging federal agencies to transition to ISO metadata.	http://www.fgdc.gov/metadata/geospatial-metadata-standards/	http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf
	ISO 19115	An internationally-adopted schema for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. Sponsored by the International Standards Organisation, the first edition of ISO 19115 was published in 2003. It has since been split into parts: ISO 19115-1:2014 contains the fundamentals of the standard; ISO 19115-2:2009 contains extensions for imagery and gridded data; and ISO/TS 19115-3:2016 provides an XML schema implementation for the fundamental concepts compatible with ISO/TS 19138:2007 (Geographic Metadata XML, or GMD).	http://www.iso.org/iso/catalogue_detail.htm?csnumber=53798	http://standards.iso.org/iso/19115/-3/
Plant pathology	*OME-XML (Open Microscopy Environment XML)*	OME-XML is a vendor-neutral file format for biological image data, with an emphasis on metadata supporting light microscopy. It can be used as a data file format in its own right, or as a way of encoding metadata within a TIFF or BigTIFF file (for which purpose there is the OME-TIFF specification). The standard is maintained by the Open Microscopy Environment Consortium, and was last updated in June 2012.	http://www.openmicroscopy.org/site/support/file-formats	http://www.openmicroscopy.org/Schemas/

Subject	Standard	Content	Website	Specification
Plant physiology	*OME-XML (Open Microscopy Environment XML)*	OME-XML is a vendor-neutral file format for biological image data, with an emphasis on metadata supporting light microscopy. It can be used as a data file format in its own right, or as a way of encoding metadata within a TIFF or BigTIFF file (for which purpose there is the OME-TIFF specification). The standard is maintained by the Open Microscopy Environment Consortium, and was last updated in June 2012.	http://www.openmicroscopy.org/site/support/file-formats	http://www.openmicroscopy.org/Schemas/
Proteomics	*MIBBI (Minimum Information for Biological and Biomedical Investigations)*	A common portal to a group of nearly 40 checklists of Minimum Information for various biological disciplines. The MIBBI Foundry is developing a cross-analysis of these guidelines to create an intercompatible, extensible community of standards. The concept was realized initially through the joint efforts of the Proteomics Standards Initiative, the Genomic Standards Consortium and the MGED RSBI Working Groups. The latest project to register with MIBBI is the MIABie guidelines for reporting biofilm research, as of January 2012.	http://mibbi.sourceforge.net/portal.shtml	http://mibbi.sourceforge.net/foundry.shtml
Remote Sensing	*Observations and Measurements*	This encoding is an essential dependency for the OGC Sensor Observation Service (SOS) Interface Standard. More specifically, this standard defines XML schemas for observations, and for features involved in sampling when making observations. These provide document models for the exchange of information describing observation acts and their results, both within and between different scientific and technical communities.	http://www.opengeospatial.org/standards/om	http://portal.opengeospatial.org/files/?artifact_id=41579
	UKEOF	A metadata standard for describing environmental monitoring activities, programmes, networks and facilities published by the UK Environmental Observation Framework (UKEOF).	http://www.ukeof.org.uk/	http://schema.ukeof.org.uk/

Subject	Standard	Content	Website	Specification
Soil Science	*UKEOF*	A metadata standard for describing environmental monitoring activities, programmes, networks and facilities published by the UK Environmental Observation Framework (UKEOF).	http://www.ukeof.org.uk/	http://schema.ukeof.org.uk/
Statistics	*DDI (Data Documentation Initiative)*	A widely used, international standard for describing data from the social, behavioral, and economic sciences. Two versions of the standard are currently maintained in parallel:DDI Codebook (or DDI version 2) is the simpler of the two, and intended for documenting simple survey data for exchange or archiving. Version 2.5 was released in January 2014.DDI Lifecycle (or DDI version 3) is richer and may be used to document datasets at each stage of their lifecycle from conceptualization through to publication and reuse. It is modular and extensible. Version 3.2 was published in March 2014.Both versions are XML-based and defined using XML Schemas. They were developed and are maintained by the DDI Alliance.	http://www.ddialliance.org/	http://www.ddialliance.org/Specification/
	RDF Data Cube Vocabulary	The standard provides a means to publish multi-dimensional data, such as statistics, on the web in such a way that it can be linked to related data sets and concepts using the W3C RDF (Resource Description Framework) standard. The model underpinning the Data Cube vocabulary is compatible with the cube model that underlies SDMX (Statistical Data and Metadata eXchange), an ISO standard for exchanging and sharing statistical data and metadata among organizations.	http://www.w3.org/TR/vocab-data-cube/	http://www.w3.org/TR/vocab-data-cube/
	SDMX (Statistical Data and Metadata Exchange)	A set of common technical and statistical standards and guidelines to be used for the efficient exchange and sharing of statistical data and metadata. Sponsoring institutions include BIS, ECB, EUROSTAT, IMF, OECD, UN, and the World Bank. Technical Specification 2.1 was amended in May 2012.	http://sdmx.org	http://sdmx.org/?page_id=5008

Subject	Standard	Content	Website	Specification
	Darwin Core	A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries. Sponsored by Biodiversity Information Standards (TWDG), the current standard was last modified in October 2009.	http://rs.tdwg.org/dwc/index.htm	http://www.tdwg.org/standards/450
Topography	*FGDC/CSDGM (Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata)*	A widely-used, but no longer current standard defining the information content for a set of digital geospatial data required by the US Federal Government.CSDGM was sponsored by the US Federal Geographic Data Committee. However, in September 2010 the FGDC endorsed ISO 19115 and began encouraging federal agencies to transition to ISO metadata.	http://www.fgdc.gov/metadata/geospatial-metadata-standards/	http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf
	ISO 19115	An internationally-adopted schema for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. Sponsored by the International Standards Organisation, the first edition of ISO 19115 was published in 2003. It has since been split into parts: ISO 19115-1:2014 contains the fundamentals of the standard; ISO 19115-2:2009 contains extensions for imagery and gridded data; and ISO/TS 19115-3:2016 provides an XML schema implementation for the fundamental concepts compatible with ISO/TS 19138:2007 (Geographic Metadata XML, or GMD).	http://www.iso.org/iso/catalogue_detail.htm?csnumber=53798	http://standards.iso.org/iso/19115/-3/
Zoology	*ABCD (Access to Biological Collection Data)*	The Access to Biological Collections Data (ABCD) Schema is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data). The ABCD Schema attempts to be comprehensive and highly	http://wiki.tdwg.org/ABCD	http://www.tdwg.org/standards/115/

Subject	Standard	Content	Website	Specification
		structured, supporting data from a wide variety of databases. It is compatible with several existing data standards. Parallel structures exist so that either (or both) atomised data and free-text can be accommodated. Sponsored by Biodiversity Information Standards TDWG - the Taxonomic Databases Working Group, the current specification was last modified in 2007.		
	Darwin Core	A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries. Sponsored by Biodiversity Information Standards (TWDG), the current standard was last modified in October 2009.	http://rs.tdwg.org/dwc/index.htm	http://www.tdwg.org/standards/450
	EML (Ecological Metadata Language)	Ecological Metadata Language (EML) is a metadata specification particularly developed for the ecology discipline. It is based on prior work done by the Ecological Society of America and associated efforts (Michener et al., 1997, Ecological Applications). Sponsored by ecoinformatics.org, EML Version 2.1.1 was released in 2011.	http://knb.ecoinformatics.org/software/eml/	http://knb.ecoinformatics.org/software/eml/eml-2.1.1/index.html

6. Metadata quality control

The creation of metadata automatically or by information originators who are not familiar with cataloging, indexing, or vocabulary control can create quality problems. Mandatory elements may be missing or used incorrectly. Schema syntax may have errors that prevent the metadata from being processed correctly. Metadata content terminology may be inconsistent, making it difficult to locate relevant information.

The National Information Standards Organization[30] addresses the metadata quality problem in the report *A Framework of Guidance for Building Good Digital Collections* for metadata created automatically (by machine) or manually (by non-cataloging professionals who are unfamiliar with cataloging, indexing or vocabulary control), providing a framework of guidance for building robust digital collections. NISO presents the six principles of what is termed *good* metadata:

- conforms to community standards;
- supports interoperability;
- uses authority control and content standards;
- includes a clear statement of the conditions and terms of use;
- supports long-term curatorship and preservation;
- should have the qualities of good objects, including authority, authenticity, archivability, persistence, and unique identification.

In other words, the quality of metadata should reflect the degree to which it performs the core bibliographic functions of discovery, use, provenance, currency, authenticity, and administration, its principal purpose being that of finding, identifying, selecting, and obtaining items.

The interoperability and exchange of metadata is further facilitated by metadata crosswalks. A crosswalk is a mapping of the elements, semantics, and syntax from one metadata scheme to those of another. A crosswalk allows metadata created by one community to be used by another group that employs a different metadata standard. The degree to which these crosswalks are successful at the individual record level depends on the similarity of the two schemes, the granularity of the elements in the target scheme compared to that of the source, and the compatibility of the content rules used to fill the elements of each scheme. Crosswalks are important for virtual collections where resources are drawn from a variety of sources and are expected to act as a whole, perhaps with a single search engine applied. While these crosswalks are key, they are also labor intensive to develop and maintain. The mapping of schemes with fewer elements (less granularity) to those with more elements (more granularity) is problematic.

As repositories grow (through automatic metadata generation or resource decomposition) and merge (through metadata harvesting), quality issues become more apparent. New techniques that take advantage of the ability of computers to perform repetitive calculations have been developed to assure a minimum level of quality. Table 14 presents two general approaches that can be outlined from the work on metadata quality evaluation for digital repositories.

[30] http://www.niso.org/publications/rp/framework3.pdf

Table 14. Two metadata quality evaluation approaches

Approach	Description
Manual Quality Evaluation	Manually review a statistically significant sample of metadata instances against a predefined set of quality parameters. Human evaluations are averaged and an estimation of metadata quality in the repository is obtained.
Simple Statistical Quality Evaluation	Automatically collect statistical information from all the metadata instances in the repository to obtain an estimation of their quality.

Two ideal characteristics of metadata quality for the fast-growing repositories should be: to be automatically calculated for each metadata instance inserted in the repository (scalability) and to provide a useful measurement of the quality (meaningfulness). In this respect, manual evaluations are meaningful but not scalable and simple statistics are scalable, but are not meaningful. The tasks metadata should enable in a digital repository are to help the user to find, identify, select and obtain the desired resources. The quality of the metadata should be directly proportional to how much it facilitates those tasks. In order to reduce subjectivity in the assessment of information quality, several quality evaluation frameworks have been developed. These frameworks define parameters that indicate whether information should be considered of high quality. In particular, *Statistics Canada's Quality Assurance Framework* presents six dimensions of information quality: relevance, accuracy, timeliness, accessibility, interpretability, and coherence. These six dimensions have been further refined to completeness, accuracy, provenance, conformance to expectation, logical consistency, coherence, timeliness, and accessibility in the chapter "The Continuum of Metadata Quality: Defining, Expressing, Exploiting," by Thomas R. Bruce and Diane Hillmann[31].

Another, more complete, Information Quality (IQ) framework was proposed by Stvilia and his collaborators[32]. Their framework consists of 21 quality dimensions comprising three categories of intrinsic, relational and reputational IQs. Assessment of data quality by intrinsic IQ dimensions can be examined through attributes of the objects by measuring conformance to a given standard. The criteria for the intrinsic IQ are accuracy/validity, cohesiveness, complexity, semantic consistency, structural consistency, currency, informativeness, naturalness, and precision. The relational IQ dimensions concern relationships between an information object and its usage context. The criteria for the relational/contextual IQ are accuracy, completeness, complexity, latency/speed, naturalness, informativeness, relevance (aboutness), precision, security, verifiability, and volatility. The reputational IQ concerns a criterion of authority centering on reputation of an information object in a given community. An overlapping relation between the frameworks of Bruce and Hillmann and Stvilia et al. can be summarized by the diagram in Figure 8 according to Park[33].

[31] Thomas R. Bruce and Diane Hillmann, "The Continuum of Metadata Quality: Defining, Expressing, Exploiting," in Metadata in Practice, eds. D. Hillmann and E. L. Westbrooks (Chicago: American Library Association, 2004)

[32] B. Stvilia, L. Gasser, M. Twidale, S. Shreeves, and T.Cole, "Metadata Quality for Federated Collections" (International Conference on Information Quality, ICIQ 2004, Cambridge, MA, 2004)

[33] Jung-Ran Park (2009) Metadata Quality in Digital Repositories: A Survey of the Current State of the Art, Cataloging & Classification Quarterly, 47:3-4, 213-228

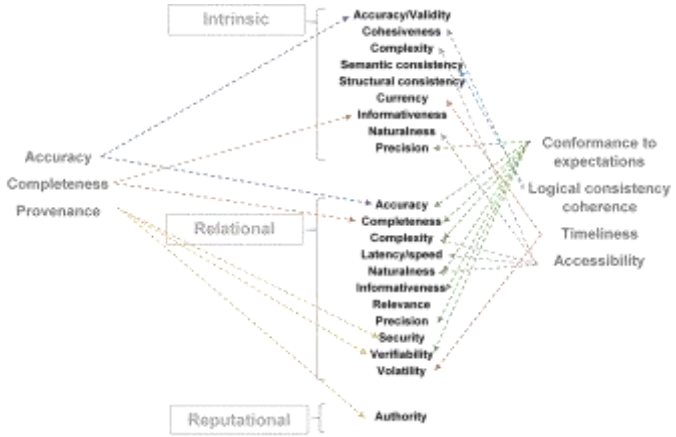

Figure 8. Mapping between Bruce & Hillmann and Stvillia quality frameworks

Due to its broad coverage (it captures other frameworks quality dimensions), compactness and ease of understanding by human reviewers, the metadata quality framework proposed by Bruce and Hillmann is suitable for DANUBIUS-RI purposes. The seven parameters used to measure the quality of metadata in this framework are described bellow:

- *Completeness*: A metadata element should describe the resource as fully as possible also taking into account the economical aspect. The metadata fields should be filled in for the majority of the resource population in order to make them useful.
- *Accuracy*: The information provided about the resource in the metadata instance should be as correct as possible. The desciption of the metadata instance should adhere to appropriate syntax rules and vocabularies.
- *Conformance to Expectations*: The degree to which metadata meets the requirements of a given community of users for a given task. If the information stored in the metadata helps a community to find, identify, select and obtain resources it could be considered to conform to the expectations of the community.
- *Logical Consitency and Coherence*: Metadata should be consistent with standard definitions and concepts used in the domain and also have internal coherence, that is that all the fields should describe the same resource.
- *Accessibility*: End users should be able to read and understand the metadata, otherwise the metadata has no value.
- *Timeliness*: Metadata changes keep up with resource changes, and up-to-date metadata is available for distribution together with a resource.
- *Provenance*: The circumstances under which the metadata instance was created, who created it, the level of expertise of the indexer, what methodologies were followed at indexing time and what subsequent transformations the metadata has passed through.

Based on the above framework, several quality metrics intended to describe the quality parameters have been proposed in the literature[34]. These metrics can be used to build tools for digital repositories and provide scalable and meaningful metadata quality assurance. Table 15 presents the formulas and the description of the most common metrics:

[34] Ochoa, X., Duval, E.: Automatic evaluation of metadata quality in digital repositories. International Journal on Digital Libraries 10(2-3), 67–91 (2009)

Table 15. A set of quality metrics for Bruce and Hillmann framework

Framework dimension	Quality metric	Description
Completeness	$Q_{comp} = \frac{\sum_{i=1}^{N} P(i)}{N}$ - $P(i)$ is 1 if the i[th] field has no-null value and 0 otherwise; - N is the number of fields	A basic completeness metric that counts the number of fields in each metadata instance that contain a no-null value (a measure of how much information is available about the resource).
	$Q_{comp} = \frac{\sum_{i=1}^{N} a_i P(i)}{\sum_{i=1}^{N} a_i}$ - a_i is the relative importance of the i[th] field	The weighted completeness metric measure the completeness of metadata instance against the actual needs of a community. It accounts for metadata elements that are more/less relevant assigning a weighting factor that represents the importance of the field.
Accuracy	$Q_{accu} = \frac{\sum_{i=1}^{N} tf(resource_i) \times tf(metadata_i)}{\sqrt{\sum_{i=1}^{N} tf(resource_i)^2 \times \sum_{i=1}^{N} tf(metadata_i)^2}}$ - $tf(resource_i)$ and $tf(metadata_i)$ are the relative frequencies of the i[th] word in the text content of the resource - N is the total number of words in both texts	The accuracy of a metadata value can be viewed as the semantic distance between the information a user could extract from the metadata and the information the same user could obtain from the resource itself. The shorter the distance the higher the accuracy. A document term

Framework dimension	Quality metric	Description
		frequency matrix is constructed, where the rows are the resource and the metadata and the columns are the words in the original resource. The intersection of rows and columns is the frequecy of the corresponding word appearance. One problem with this metric is given by the availability of the resource itself.
	$Q_{cinfo} = \frac{\sum_{i=1}^{N} infoContent(field_i)}{N}$, $infoContent(cat\_field) = -\log(f(value))$, - $f(value)$ is the relative frequency of value in categorical field for all the current instances - N is the number of categorical fields	The information content of categorical fields is computed by averaging the entropy (as a measure of the information content) of the categorical fields.
Conformance to Expectations	$Q_{tinfo} = \log\left(\sum_{i=1}^{N} infoContent(field_i)\right)$, $infoContent(freetext) = \sum_{i=1}^{N} tf(word_i) \times \log\left(\frac{1}{df(word_i)}\right)$ - $tf(word_i)$ is the term frequency of the i[th] word - $df(word_i)$ is the document frequency of the i[th] word - N is the number of words in the field	For the free text fields the information content calculation is based on the Term Frequency Inverse Document Frequency value. The importance of a word in a document is directly proportional to how frequently this word appears in the document and inversely proportional to how frequently documents contain this word.

Framework dimension	Quality metric	Description
		The Q_{tinfo} metric tries to estimate the total information content of the instance.
	$$Q_{cons} = 1 - \frac{\sum_{i=1}^{N} brokeRule_i}{N},$$ $$brokeRule_i = \begin{cases} 0, & \text{if instance complies with i-th rule} \\ 1, & \text{otherwise} \end{cases}$$ - N is the number of rules in the metadata standard	Metadata should be consisitent with standard definitions and concepts used in the domain. The consistency of the metadata instance should be inversely proportional to the number of problems found in the instance.
Logical Consitency and Coherence	$$Q_{coh} = \frac{\sum_{i}^{N}\sum_{j}^{N} \begin{cases} d(field_i, field_j), & i < j \\ 0, & otherwise \end{cases}}{\frac{N(N-1)}{2}},$$ $$d(f_1, f_2) = \frac{\sum_{i=1}^{N} tfidf_{i,f_1} \times tfidf_{i,f_2}}{\sqrt{\sum_{i=1}^{N} tfidf_{i,f_1}{}^2 \times \sum_{i=1}^{N} tfidf_{i,f_2}{}^2}},$$ - $tfidf_{i,field}$ is the term frequency inverse document frequency of the i[th] word in the textual field f - N is the number of textual fields that describe the object	The coherence of the instance is related to the degree to which all fields describe the same object in a similar way. A Term Frequency Inverse Document Frequency matrix is computed in order to reflect the importance of words to the corresponding documents.
Accessibility	$$Q_{link} = \frac{links(inst_k)}{\max_{i=1,N} links(inst_i)},$$ - $links(inst)$ represent the number of pointers to or from the metadata instance - N is the number of resources	Accessibility implies the level to which a metadata instance can be found and later understood. A network model is proposed such that a link is implicit (objects of the same authors) or explicit (is

Framework dimension	Quality metric	Description
		related to or is version of). The logical accessibility can be calculated as the number of links from the nodes to other nodes.
Timeliness	$$Q_{time} = \frac{Q_{curr_{t_2}} - Q_{curr_{t1}}}{Q_{curr_{t1}}(t_2 - t_1)},$$ $$Q_{curr} = Q_{avg} = \frac{\sum_{i=1}^{N} \frac{Q_i - \min Q_j}{\max Q_j - \min Q_j}}{N},$$ - Q_i is the value of the i[th] quality metric (for example Q_{comp} and Q_{tinfo}) - $\min Q_j$ and $\max Q_j$ are the minimum and maximum value of the j[th] metric for all the instances in the repository - N is the number of metrics considered	The timeliness in digital repositories mainly relates to the degree to which a metadata instance remains current. An estimation of the timeliness of an instance could be obtained by measuring the rate of change of the instantaneous currency over a period of time. In other terms, the timeliness of an instance will be equal to its change of average quality per unit of time and can be used to estimate how much better the instance will be after a predefined period.
Provenance	$$Q_{prov} = \frac{\sum_{i=1}^{N} Q_{avg_i}}{N},$$ - Q_{avg_i} is the average quality of the i[th] instance contributed by the source S - N is the total number of instances produced by S	Provenance quality measures the trust that a given community has in the source of the metadata instance. The Q_{prov} metric can be calculated once the other quality metrics has been calculated and assigned to each instance. It is calculated as the

Framework dimension	Quality metric	Description
		average (of all instances produced by the source) of the Average Quality for each instance.

7. References

1. Alemu G, Stevens B (2015) Existing standards-based metadata approaches and principles. 11–28. doi: 10.1016/B978-0-08-100385-5.00002-X
2. Bailo D, Ulbricht D, Nayembil ML, Trani L, Spinuso A, Jeffery KG (2017) Mapping Solid Earth Data and Research Infrastructures to CERIF. Procedia Computer Science 106:112–121. doi: 10.1016/j.procs.2017.03.043
3. Bargmeyer BE, Gillman DW (2000) Metadata standards and Metadata registries: an overview. Vasa 10.
4. Biagi L de, Saccone M, Trufelli L, Puccinelli R (2012) Research product repositories: Strategies for data and metadata quality control. Grey Journal 8:83–94. doi: 10.2777/96979
5. Brand A, Daly F, Meyers B (2003) Metadata Demystified. 1–19.
6. Bui Y, Park J-R (2006) An assessment of metadata quality: A case study of the National Science Digital Library Metadata Repository. Proceedings of CAISACSI 2006 Information Science Revisited Approaches to Innovation 13.
7. Cleary J, Holmes G, Cunningham SJ, Witten IH (1996) MetaData for database mining. Proc IEEE Metadata …
8. Cole TW (2002) Creating a framework of guidance for building good digital collections. doi: 10.5210/fm.v7i5.955
9. Cortês C (2005) Humanities, Computers and Cultural Heritage : Proceedings of the XVI international conference of the Association for History and Computing. 12:308.
10. Díaz P, Masó J, Sevillano E, Ninyerola M, Zabala A, Serral I, Pons X (2012) Analysis of quality metadata in the GEOSS Clearinghouse. International Journal of Spatial Data Infrastructures Research 7:352–377. doi: 10.2902/1725-0463.2012.07.art17
11. Drafting Team Metadata and European Commission Joint Research Centre (2013) INSPIRE Metadata Implementing Rules : Technical Guidelines based on EN ISO 19115 and EN ISO 19119 - V.1.3. 99. doi: MD_IR_and_ISO_20090218
12. European Commission (2017) Technical Guidelines for implementing dataset and service metadata based on ISO/TS 19139:2007. 178.
13. Fox M, Rinehart R (2005) Descriptive Metadata Guidelines fro RLG Cutlrual Materials. 64.

14. Franks P, Ph D, Kunde N (2006) Records managers must be involved in the development and design of metadata structures to ensure that digital records are captured, maintained, retained, preserved, or destroyed in accordance with their organization's recordkeeping requirements. The Information Management Journal 13:235–246.

15. Froeschl KA, Grossmann W, Del Vecchio V (2003) The Concept of Statistical Metadata. MetaNet (IST-1999-29093) Work Group 2:

16. Gilliland A, Rouche N, Lindberg L, Evans J (2005) Towards a 21st century metadata infrastructure supporting the creation, preservation and use of trustworthy records: Developing the InterPARES 2 metadata schema registry. Archival Science 5:43–78. doi: 10.1007/s10502-005-9000-4

17. Habermann T International Metadata Standards and Enterprise Data Quality Metadata Systems. DataOne

18. Ho K (2003) Data Documentation - MCHP. 2002–2003.

19. Jeffery KG, Koskela R (2015) RDA: The Importance of Metadata. ERCIM News 100. Special Theme: Scientific Data Sharing and Re-use

20. Klas W (2008) Metadata for semantic and social applications DC- 2008 Berlin. doi: 10.17875/gup2008-446

21. Kunze J, Baker T (2007) The Dublin Core Metadata Element Set. doi: 10.17487/rfc5013

22. Kübler S, Skala W, Voisard A (1998) the Design and Development of a Geologic Hypermap Prototype (1). Symposium A Quarterly Journal In Modern Foreign Literatures 32:336–343.

23. Löbe M, Knuth M, Mücke R (2009) TIM: A semantic web application for the specification of metadata items in clinical research. CEUR Workshop Proceedings 559:

24. Nations U (1995) Guidelines for the modeling of statistical data and metadata. Design

25. Ochoa X, Duval E (2009) Automatic evaluation of metadata quality in digital repositories. International Journal on Digital Libraries 10:67–91. doi: 10.1007/s00799-009-0054-4

26. Park J-R (2009) Metadata Quality in Digital Repositories: A Survey of the Current State of the Art. Cataloging & Classification Quarterly 47:213–228. doi: 10.1080/01639370902737240

27. Park, Jung-ranTosaka, YujiMaszaros, SusanLu C (2010) From Metadata Creation to Metadata Quality Control: Continuing Education Needs Among Cataloging and Metadata Professionals. Journal ofJournal of Education for Library & Information Science 51:p158–176.

28. Premis Editorial Committee (2008) Data Dictionary section from PREMIS Data Dictionary for Preservation Metadata. Preservation 1–184. doi: 10.1515/9783110197303.toc

29. Strasser C (2015) Research Data Management. 27.

30. Stvilia B, Gasser L, Twidale MB, Smith LC (2007) A framework for information quality assessment. Journal of the American Society for Information Science and Technology 58:1720–1733. doi: 10.1002/asi.20652

31. Turvey MR (2005) Metadata in Practice (review). 5:137–138. doi: 10.1353/pla.2005.0012

32. Workflow II, Quality III, Accuracy II, Conformance III, Consistency IVL, Timeliness VI, Provenance VII, Tools IV, Related VI (2015) A Metadata Quality Assurance Framework. 1–11.

Appendix

Appendix A. Premis Data Dictionary semantic units

This is a hierarchical listing of the semantic units included in the Data Dictionary. The information in parentheses following each semantic unit indicates whether the element is Mandatory (M) or Optional (O), followed by whether it is Repeatable (R) or Not repeatable (NR).

Table 16. PREMIS Data Dictionary

Entities	List of Semantic Units
OBJECT	1.1 objectIdentifier (M, R) 1.1.1 objectIdentifierType (M, NR) 1.1.2 objectIdentifierValue (M, NR) 1.2 objectCategory (M, NR) 1.3 preservationLevel (O, R) [Intellectual Entity, Representation, File] 1.3.1 preservationLevelType (O, NR) [Intellectual Entity, Representation, File] 1.3.2 preservationLevelValue (M, NR) [Intellectual Entity, Representation, File] 1.3.3 preservationLevelRole (O, NR) [Intellectual Entity, Representation, File] 1.3.4 preservationLevelRationale (O, R) [Intellectual Entity, Representation, File] 1.3.5 preservationLevelDateAssigned (O, NR) [Intellectual Entity, Representation, File] 1.4 significantProperties (O, R) 1.4.1 significantPropertiesType (O, NR) 1.4.2 significantPropertiesValue (O, NR) 1.4.3 significantPropertiesExtension (O, R) 1.5 objectCharacteristics (M, R) [File, Bitstream] 1.5.1 compositionLevel (O, NR) [File, Bitstream] 1.5.2 fixity (O, R) [File, Bitstream] 1.5.2.1 messageDigestAlgorithm (M, NR) [File, Bitstream] 1.5.2.2 messageDigest (M, NR) [File, Bitstream]

Entities	List of Semantic Units
	1.5.2.3 messageDigestOriginator (O, NR) [File, Bitstream] 1.5.3 size (O, NR) [File, Bitstream] 1.5.4 format (M, R) [File, Bitstream] 1.5.4.1 formatDesignation (O, NR) [File, Bitstream] 1.5.4.1.1 formatName (M, NR) [File, Bitstream] 1.5.4.1.2 formatVersion (O, NR) [File, Bitstream] 1.5.4.2 formatRegistry (O, NR) [File, Bitstream] 1.5.4.2.1 formatRegistryName (M, NR) [File, Bitstream] 1.5.4.2.2 formatRegistryKey (M, NR) [File, Bitstream] 1.5.4.2.3 formatRegistryRole (O, NR) [File, Bitstream] 1.5.4.3 formatNote (O, R) [File, Bitstream] 1.5.5 creatingApplication (O, R) [File, Bitstream] 1.5.5.1 creatingApplicationName (O, NR) [File, Bitstream] 1.5.5.2 creatingApplicationVersion (O, NR) [File, Bitstream] 1.5.5.3 dateCreatedByApplication (O, NR) [File, Bitstream] 1.5.5.4 creatingApplicationExtension (O, R) [File, Bitstream] 1.5.6 inhibitors (O, R) [File, Bitstream] 1.5.6.1 inhibitorType (M, NR) [File, Bitstream] 1.5.6.2 inhibitorTarget (O, R) [File, Bitstream] 1.5.6.3 inhibitorKey (O, NR) [File, Bitstream] 1.5.7 objectCharacteristicsExtension (O, R) [File, Bitstream] 1.6 originalName (O, NR) [Intellectual Entity, Representation, File] 1.7 storage (O, R) [Representation, File, Bitstream] 1.7.1 contentLocation (O, NR) [Representation, File, Bitstream] 1.7.1.1 contentLocationType (M, NR) [Representation, File, Bitstream] 1.7.1.2 contentLocationValue (M, NR) [Representation, File, Bitstream] 1.7.2 storageMedium (O, NR) [Representation, File, Bitstream] 1.8 signatureInformation (O, R) [File, Bitstream] 1.8.1 signature (O, R) [File, Bitstream] 1.8.1.1 signatureEncoding (M, NR) [File, Bitstream] 1.8.1.2 signer (O, NR) [File, Bitstream] 1.8.1.3 signatureMethod (M, NR) [File, Bitstream] 1.8.1.4 signatureValue (M, NR) [File, Bitstream] 1.8.1.5 signatureValidationRules (M, NR) [File, Bitstream] 1.8.1.6 signatureProperties (O, R) [File, Bitstream] 1.8.1.7 keyInformation (O, NR) [File, Bitstream] 1.8.2 signatureInformationExtension (O, R) [File, Bitstream] 1.9 environmentFunction (O, R) [Intellectual Entity of type environment] 1.9.1 environmentFunctionType (M, NR) [Intellectual Entity of type environment] 1.9.2 environmentFunctionLevel (M, NR) [Intellectual Entity of type environment] 1.10 environmentDesignation (O, R) [Intellectual Entity of type environment] 1.10.1 environmentName (M, NR) [Intellectual Entity of type environment] 1.10.2 environmentVersion (O, NR) [Intellectual Entity of type environment] 1.10.3 environmentOrigin (O, NR) [Intellectual Entity of type environment] 1.10.4 environmentDesignationNote (O, R) [Intellectual Entity of type environment] 1.10.5 environmentDesignationExtension (O, R) [Intellectual Entity of type environment] 1.11 environmentRegistry (O, R) [Intellectual Entity of type environment] 1.11.1 environmentRegistryName (M, NR) [Intellectual Entity of type environment] 1.11.2 environmentRegistryKey (M, NR) [Intellectual Entity of type environment]

Entities	List of Semantic Units
	1.11.3 environmentRegistryRole (O, NR) [Intellectual Entity of type environment] 1.12 environmentExtension (O, R) [Intellectual Entity of type environment] 1.13 relationship (O, R) 1.13.1 relationshipType (M, NR) 1.13.2 relationshipSubType (M, NR) 1.13.3 relatedObjectIdentifier (M, R) 1.13.3.1 relatedObjectIdentifierType (M, NR) 1.13.3.2 relatedObjectIdentifierValue (M, NR) 1.13.3.3 relatedObjectSequence (O, NR) 1.13.4 relatedEventIdentifier (O, R) 1.13.4.1 relatedEventIdentifierType (M, NR) 1.13.4.2 relatedEventIdentifierValue (M, NR) 1.13.4.3 relatedEventSequence (O, NR) 1.13.5 relatedEnvironmentPurpose (O, R) [Representation, File, Bitstream] 1.13.6 relatedEnvironmentCharacteristic (O, NR) [Representation, File, Bitstream] 1.14 linkingEventIdentifier (O, R) 1.14.1 linkingEventIdentifierType (M, NR) 1.14.2 linkingEventIdentifierValue (M, NR) 1.15 linkingRightsStatementIdentifier (O, R) 1.15.1 linkingRightsStatementIdentifierType (M, NR) 1.15.2 linkingRightsStatementIdentifierValue (M, NR)
EVENT	2.1 eventIdentifier (M, NR) 2.1.1 eventIdentifierType (M, NR) 2.1.2 eventIdentifierValue (M, NR) 2.2 eventType (M, NR) 2.3 eventDateTime (M, NR) 2.4 eventDetailInformation (O, R) 2.4.1 eventDetail (O, NR) 2.4.2 eventDetailExtension (O, R) 2.5 eventOutcomeInformation (O, R) 2.5.1 eventOutcome (O, NR) 2.5.2 eventOutcomeDetail (O, R) 2.5.2.1 eventOutcomeDetailNote (O, NR) 2.5.2.2 eventOutcomeDetailExtension (O, R) 2.6 linkingAgentIdentifier (O, R) 2.6.1 linkingAgentIdentifierType (M, NR) 2.6.2 linkingAgentIdentifierValue (M, NR) 2.6.3 linkingAgentRole (O, R) 2.7 linkingObjectIdentifier (O, R) 2.7.1 linkingObjectIdentifierType (M, NR) 2.7.2 linkingObjectIdentifierValue (M, NR) 2.7.3 linkingObjectRole (O, R)
AGENT	3.1 agentIdentifier (M, R) 3.1.1 agentIdentifierType (M, NR) 3.1.2 agentIdentifierValue (M, NR) 3.2 agentName (O, R) 3.3 agentType (O, NR) 3.4 agentVersion (O, NR) 3.5 agentNote (O, R) 3.6 agentExtension (O, R) 3.7 linkingEventIdentifier (O, R) 3.7.1 linkingEventIdentifierType (M, NR)

Entities	List of Semantic Units
	3.7.2 linkingEventIdentifierValue (M, NR) 3.8 linkingRightsStatementIdentifier (O, R) 3.8.1 linkingRightsStatementIdentifierType (M, NR) 3.8.2 linkingRightsStatementIdentifierValue (M, NR) 3.9 linkingEnvironmentIdentifier (O, R) 3.9.1 linkingEnvironmentIdentifierType (M, NR) 3.9.2 linkingEnvironmentIdentifierValue (M, NR) 3.9.3 linkingEnvironmentRole (O, R)
RIGHTS	4.1 rightsStatement (O, R) 4.1.1 rightsStatementIdentifier (M, NR) 4.1.1.1 rightsStatementIdentifierType (M, NR) 4.1.1.2 rightsStatementIdentifierValue (M, NR) 4.1.2 rightsBasis (M, NR) 4.1.3 copyrightInformation (O, NR) 4.1.3.1 copyrightStatus (M, NR) 4.1.3.2 copyrightJurisdiction (M, NR) 4.1.3.3 copyrightStatusDeterminationDate (O, NR) 4.1.3.4 copyrightNote (O, R) 4.1.3.5 copyrightDocumentationIdentifier (O, R) 4.1.3.5.1 copyrightDocumentationIdentifierType (M, NR) 4.1.3.5.2 copyrightDocumentationIdentifierValue (M, NR) 4.1.3.5.3 copyrightDocumentationRole (O, NR) 4.1.3.6 copyrightApplicableDates (O, NR) 4.1.3.6.1 startDate (O, NR) 4.1.3.6.2 endDate (O, NR) 4.1.4 licenseInformation (O, NR) 4.1.4.1 licenseDocumentationIdentifier (O, R) 4.1.4.1.1 licenseDocumentationIdentifierType (M, NR) 4.1.4.1.2 licenseDocumentationIdentifierValue (M, NR) 4.1.4.1.3 licenseDocumentationRole (O, NR) 4.1.4.2 licenseTerms (O, NR) 4.1.4.3 licenseNote (O, R) 4.1.4.4 licenseApplicableDates (O, NR) 4.1.4.4.1 startDate (O, NR) 4.1.4.4.2 endDate (O, NR) 4.1.5 statuteInformation (O, R) 4.1.5.1 statuteJurisdiction (M, NR) 4.1.5.2 statuteCitation (M, NR) 4.1.5.3 statuteInformationDeterminationDate (O, NR) 4.1.5.4 statuteNote (O, R) 4.1.5.5 statuteDocumentationIdentifier (O, R) 4.1.5.5.1 statuteDocumentationIdentifierType (M, NR) 4.1.5.5.2 statuteDocumentationIdentifierValue (M, NR) 4.1.5.5.3 statuteDocumentationRole (O, NR) 4.1.5.6 statuteApplicableDates (O, NR) 4.1.5.6.1 startDate (O, NR) 4.1.5.6.2 endDate (O, NR) 4.1.6 otherRightsInformation (O, NR) 4.1.6.1 otherRightsDocumentationIdentifier (O, R) 4.1.6.1.1 otherRightsDocumentationIdentifierType (M, NR) 4.1.6.1.2 otherRightsDocumentationIdentifierValue (M, NR) 4.1.6.1.3 otherRightsDocumentationRole (O, NR) 4.1.6.2 otherRightsBasis (M, NR) 4.1.6.3 otherRightsApplicableDates (O, NR) 4.1.6.3.1 startDate (O, NR)

Entities	List of Semantic Units
	4.1.6.3.2 endDate (O, NR) 4.1.6.4 otherRightsNote (O, R) 4.1.7 rightsGranted (O, R) 4.1.7.1 act (M, NR) 4.1.7.2 restriction (O, R) 4.1.7.3 termOfGrant (O, NR) 4.1.7.3.1 startDate (M, NR) 4.1.7.3.2 endDate (O, NR) 4.1.7.4 termOfRestriction (O, NR) 4.1.7.4.1 startDate (M, NR) 4.1.7.4.2 endDate (O, NR) 4.1.7.5 rightsGrantedNote (O, R) 4.1.8 linkingObjectIdentifier (O, R) 4.1.8.1 linkingObjectIdentifierType (M, NR) 4.1.8.2 linkingObjectIdentifierValue (M, NR) 4.1.8.3 linkingObjectRole (O, R) 4.1.9 linkingAgentIdentifier (O, R) 4.1.9.1 linkingAgentIdentifierType (M, NR) 4.1.9.2 linkingAgentIdentifierValue (M, NR) 4.1.9.3 linkingAgentRole (O, R) 4.2 rightsExtension (O, R)

Appendix B. EXif Schema for digital still images

Table 17. EXif elements and their description

Element	Description
exifdata	An Exif IFD data entry
tag_number	The Exif tag number
tagid	The Exif tag number with context prefix, such as IFD type or maker name
datatype	The Exif field data type, such as ascii, byte, short etc.
length	Length of an object. Could be a subProperty of other general schema.
width	Width of an object
height	Height of an object
resolution	a rational number representing a resolution. Could be a subProperty of other general schema.
meter	A length with unit of meter
mm	A length with unit of mm
seconds	a mesurement of time length with unit of second

Element	Description
date	a date information. Usually saved as YYYY:MM:DD (HH:MM:SS) format in Exif data, but represented here as W3C-DTF format
subseconds	A tag used to record fractions of seconds for a date property
geo	Geometric data such as latitude, longitude and altitude. Usually saved as rational number.
_unknown	An Exif tag whose meaning is not known
exifAttribute	A property that connects an IFD to one of its entries. Super property which integrates all Exif tags.
dateAndOrTime	An attribute relating to Date and/or Time
gpsInfo	An attribute relating to GPS information
ifdPointer	A tag that refers a child IFD
imageConfig	An attribute relating to Image Configuration
imageDataCharacter	An attribute relating to image data characteristics
imageDataStruct	An attribute relating to image data structure
interopInfo	character(0)
pictTaking	An attribute relating to Picture-Taking Conditions
pimInfo	An attribute relating to print image matching
recOffset	An attribute relating to recording offset
relatedFile	Tag Relating to Related File Information
userInfo	An attribute relating to User Information
versionInfo	An attribute relating to Version
imageWidth	Image width. The number of columns of image data, equal to the number of pixels per row. In JPEG compressed data a JPEG marker is used instead of this tag.
imageLength	Image height. The number of rows of image data. In JPEG compressed data a JPEG marker is used.
bitsPerSample	The number of bits per image component. In this standard each component of the image is 8 bits, so the value for this tag is 8. See also SamplesPerPixel. In JPEG compressed data a JPEG marker is used instead of this tag.

Element	Description
compression	The compression scheme used for the image data. When a primary image is JPEG compressed, this designation is not necessary and is omitted. When thumbnails use JPEG compression, this tag value is set to 6.
photometricInterpretation	Pixel composition. In JPEG compressed data a JPEG marker is used instead of this tag.
imageDescription	A character string giving the title of the image. It may be a comment such as “1988 company picnic” or the like. Two-byte character codes cannot be used. When a 2-byte code is necessary, the Exif Private tag UserComment is to be used.
make	Manufacturer of image input equipment
model	Model of image input equipment
stripOffsets	For each strip, the byte offset of that strip. With JPEG compressed data this designation is not needed and is omitted.
orientation	The image orientation viewed in terms of rows and columns.
samplesPerPixel	The number of components per pixel. Since this standard applies to RGB and YCbCr images, the value set for this tag is 3. In JPEG compressed data a JPEG marker is used instead of this tag.
rowsPerStrip	The number of rows per strip. This is the number of rows in the image of one strip when an image is divided into strips. With JPEG compressed data this designation is not needed and is omitted.
stripByteCounts	The total number of bytes in each strip. With JPEG compressed data this designation is not needed and is omitted.
xResolution	The number of pixels per ResolutionUnit in the ImageWidth direction. When the image resolution is unknown, 72 [dpi] is designated.
yResolution	The number of pixels per ResolutionUnit in the ImageLength direction. The same value as XResolution is designated.
planarConfiguration	Indicates whether pixel components are recorded in chunky or planar format. In JPEG compressed files a JPEG marker is used instead of this tag. If this field does not exist, the TIFF default of 1 (chunky) is assumed.

Element	Description
resolutionUnit	The unit for measuring XResolution and YResolution. The same unit is used for both XResolution and YResolution. If the image resolution in unknown, 2 (inches) is designated.
transferFunction	A transfer function for the image, described in tabular style. Normally this tag is not necessary, since color space is specified in the color space information tag (ColorSpace).
software	The name and version of the software or firmware of the camera or image input device used to generate the image.
dateTime	The date and time of image creation. In this standard it is the date and time the file was changed.
artist	Person who created the image
whitePoint	The chromaticity of the white point of the image. Normally this tag is not necessary, since color space is specified in the color space information tag (ColorSpace).
primaryChromaticities	The chromaticity of the three primary colors of the image. Normally this tag is not necessary, since color space is specified in the color space information tag (ColorSpace).
jpegInterchangeFormat	The offset to the start byte (SOI) of JPEG compressed thumbnail data. This is not used for primary image JPEG data.
jpegInterchangeFormatLength	The number of bytes of JPEG compressed thumbnail data. This is not used for primary image JPEG data.
yCbCrCoefficients	The matrix coefficients for transformation from RGB to YCbCr image data.
yCbCrSubSampling	The sampling ratio of chrominance components in relation to the luminance component. In JPEG compressed data a JPEG marker is used instead of this tag.
yCbCrPositioning	The position of chrominance components in relation to the luminance component. This field is designated only for JPEG compressed data or uncompressed YCbCr data.
referenceBlackWhite	The reference black point value and reference white point value. The color space is declared in a color space information tag, with the default being the value that gives the optimal image characteristics Interoperability these conditions.
copyright	Copyright information. In this standard the tag is used to indicate both the photographer and editor copyrights. It is the copyright notice of the person or organization claiming rights to the image.

Element	Description
exif_IFD_Pointer	A pointer to the Exif IFD, which is a set of tags for recording Exif-specific attribute information.
gpsInfo_IFD_Pointer	A pointer to the GPS IFD, which is a set of tags for recording GPS information.
exposureTime	Exposure time, given in seconds (sec).
fNumber	F number
exposureProgram	The class of the program used by the camera to set exposure when the picture is taken.
spectralSensitivity	Indicates the spectral sensitivity of each channel of the camera used. The tag value is an ASCII string compatible with the standard developed by the ASTM Technical committee.
isoSpeedRatings	Indicates the ISO Speed and ISO Latitude of the camera or input device as specified in ISO 12232.
oecf	Indicates the Opto-Electric Conversion Function (OECF) specified in ISO 14524. OECF is the relationship between the camera optical input and the image values.
exifVersion	Exif Version
dateTimeOriginal	The date and time when the original image data was generated. For a DSC the date and time the picture was taken are recorded.
dateTimeDigitized	The date and time when the image was stored as digital data. If, for example, an image was captured by DSC and at the same time the file was recorded, then the DateTimeOriginal and DateTimeDigitized will have the same contents.
componentsConfiguration	Information specific to compressed data. The channels of each component are arranged in order from the 1st component to the 4th. For uncompressed data the data arrangement is given in the PhotometricInterpretation tag. However, since PhotometricInterpretation can only express the order of Y,Cb and Cr, this tag is provided for cases when compressed data uses components other than Y, Cb, and Cr and to enable support of other sequences.
compressedBitsPerPixel	Information specific to compressed data. The compression mode used for a compressed image is indicated in unit bits per pixel.
shutterSpeedValue	Shutter speed. The unit is the APEX (Additive System of Photographic Exposure) setting

Element	Description
apertureValue	The lens aperture. The unit is the APEX value.
brightnessValue	The value of brightness. The unit is the APEX value. Ordinarily it is given in the range of -99.99 to 99.99. Note that if the numerator of the recorded value is FFFFFFFF.H, Unknown shall be indicated.
exposureBiasValue	The exposure bias. The unit is the APEX value. Ordinarily it is given in the range of -99.99 to 99.99.
maxApertureValue	The smallest F number of the lens. The unit is the APEX value. Ordinarily it is given in the range of 00.00 to 99.99, but it is not limited to this range.
subjectDistance	The distance to the subject, given in meters. Note that if the numerator of the recorded value is FFFFFFFF.H, Infinity shall be indicated; and if the numerator is 0, Distance unknown shall be indicated.
meteringMode	Metering mode, such as CenterWeightedAverage, Spot, MultiSpot,Pattern, Partial etc.
lightSource	Light source such as Daylight, Tungsten, Flash etc.
flash	The status of flash when the image was shot.
focalLength	The actual focal length of the lens, in mm. Conversion is not made to the focal length of a 35 mm film camera.
subjectArea	The location and area of the main subject in the overall scene.
makerNote	Manufacturer notes
userComment	A tag for Exif users to write keywords or comments on the image besides those in ImageDescription, and without the character code limitations of the ImageDescription tag. The character code used in the UserComment tag is identified based on an ID code in a fixed 8-byte area at the start of the tag data area.
subSecTime	DateTime subseconds
subSecTimeOriginal	DateTimeOriginal subseconds
subSecTimeDigitized	DateTimeDigitized subseconds
flashpixVersion	The Flashpix format version supported by a FPXR file. If the FPXR function supports Flashpix format Ver. 1.0, this is indicated similarly to ExifVersion by recording “0100” as 4-byte ASCII.
colorSpace	The color space information tag (ColorSpace) is always recorded as the color space specifier. Normally sRGB (=1)

Element	Description
	is used to define the color space based on the PC monitor conditions and environment.
pixelXDimension	Information specific to compressed data. When a compressed file is recorded, the valid width of the meaningful image shall be recorded in this tag, whether or not there is padding data or a restart marker. This tag should not exist in an uncompressed file.
pixelYDimension	Information specific to compressed data. When a compressed file is recorded, the valid height of the meaningful image shall be recorded in this tag, whether or not there is padding data or a restart marker. This tag should not exist in an uncompressed file. Since data padding is unnecessary in the vertical direction, the number of lines recorded in this valid image height tag will in fact be the same as that recorded in the SOF.
relatedSoundFile	Related audio file
interoperability_IFD_Pointer	A pointer to the Interoperability IFD, which is composed of tags storing the information to ensure the Interoperability
flashEnergy	The strobe energy at the time the image is captured, as measured in Beam Candle Power Seconds (BCPS).
spatialFrequencyResponse	This tag records the camera or input device spatial frequency table and SFR values in the direction of image width, image height, and diagonal direction, as specified in ISO 12233.
focalPlaneXResolution	The number of pixels in the image width (X) direction per FocalPlaneResolutionUnit on the camera focal plane.
focalPlaneYResolution	The number of pixels in the image height (Y) direction per FocalPlaneResolutionUnit on the camera focal plane.
focalPlaneResolutionUnit	The unit for measuring FocalPlaneXResolution and FocalPlaneYResolution. This value is the same as the ResolutionUnit.
subjectLocation	The location of the main subject in the scene. The value of this tag represents the pixel at the center of the main subject relative to the left edge, prior to rotation processing as per the Rotation tag. The first value indicates the X column number and second indicates the Y row number.
exposureIndex	The exposure index selected on the camera or input device at the time the image is captured.
sensingMethod	The image sensor type on the camera or input device, such as One-chip color area sensor etc.

Element	Description
fileSource	The image source. If a DSC recorded the image, this tag value of this tag always be set to 3, indicating that the image was recorded on a DSC.
sceneType	The type of scene. If a DSC recorded the image, this tag value shall always be set to 1, indicating that the image was directly photographed.
cfaPattern	The color filter array (CFA) geometric pattern of the image sensor when a one-chip color area sensor is used. It does not apply to all sensing methods.
customRendered	The use of special processing on image data, such as rendering geared to output. When special processing is performed, the reader is expected to disable or minimize any further processing.
exposureMode	the exposure mode set when the image was shot. In auto-bracketing mode, the camera shoots a series of frames of the same scene at different exposure settings.
whiteBalance	The white balance mode set when the image was shot.
digitalZoomRatio	The digital zoom ratio when the image was shot. If the numerator of the recorded value is 0, this indicates that digital zoom was not used.
focalLengthIn35mmFilm	The equivalent focal length assuming a 35mm film camera, in mm. A value of 0 means the focal length is unknown. Note that this tag differs from the FocalLength tag.
sceneCaptureType	The type of scene that was shot. It can also be used to record the mode in which the image was shot, such as Landscape, Portrait etc. Note that this differs from the scene type (SceneType) tag.
gainControl	The degree of overall image gain adjustment.
contrast	The direction of contrast processing applied by the camera when the image was shot.
saturation	The direction of saturation processing applied by the camera when the image was shot.
sharpness	The direction of sharpness processing applied by the camera when the image was shot.
deviceSettingDescription	Information on the picture-taking conditions of a particular camera model. The tag is used only to indicate the picture-taking conditions in the reader.
subjectDistanceRange	The distance to the subject, such as Macro, Close View or Distant View.

Element	Description
imageUniqueID	An identifier assigned uniquely to each image. It is recorded as an ASCII string equivalent to hexadecimal notation and 128-bit fixed length.
gpsVersionID	The version of GPSInfoIFD. The version is given as 2.2.0.0. This tag is mandatory when GPSInfo tag is present.
gpsLatitudeRef	Indicates whether the latitude is north or south latitude. The ASCII value 'N' indicates north latitude, and 'S' is south latitude.
gpsLatitude	The latitude, expressed as three values giving the degrees, minutes, and seconds, respectively.
gpsLongitudeRef	Indicates whether the longitude is east or west longitude. ASCII 'E' indicates east longitude, and 'W' is west longitude.
gpsLongitude	The longitude, expressed as three values giving the degrees, minutes, and seconds, respectively.
gpsAltitudeRef	Indicates the altitude used as the reference altitude. If the reference is sea level and the altitude is above sea level, 0 is given. If the altitude is below sea level, a value of 1 is given and the altitude is indicated as an absolute value in the GPSAltitude tag. The reference unit is meters.
gpsAltitude	The altitude based on the reference in GPSAltitudeRef. Altitude is expressed as one RATIONAL value. The reference unit is meters.
gpsTimeStamp	The time as UTC (Coordinated Universal Time). TimeStamp is expressed as three RATIONAL values giving the hour, minute, and second.
gpsSatellites	The GPS satellites used for measurements. This tag can be used to describe the number of satellites, their ID number, angle of elevation, azimuth, SNR and other information in ASCII notation. The format is not specified. If the GPS receiver is incapable of taking measurements, value of the tag shall be set to NULL.
gpsStatus	The status of the GPS receiver when the image is recorded. 'A' means measurement is in progress, and 'V' means the measurement is Interoperability.
gpsMeasureMode	The GPS measurement mode. '2' means two-dimensional measurement and '3' means three-dimensional measurement is in progress.
gpsDOP	The GPS DOP (data degree of precision). An HDOP value is written during two-dimensional measurement, and PDOP during three-dimensional measurement.

Element	Description
gpsSpeedRef	The unit used to express the GPS receiver speed of movement. ‘K’ ‘M’ and ‘N’ represents kilometers per hour, miles per hour, and knots.
gpsSpeed	The speed of GPS receiver movement.
gpsTrackRef	The reference for giving the direction of GPS receiver movement. ‘T’ denotes true direction and ‘M’ is magnetic direction.
gpsTrack	The direction of GPS receiver movement. The range of values is from 0.00 to 359.99.
gpsImgDirectionRef	The reference for giving the direction of the image when it is captured. ‘T’ denotes true direction and ‘M’ is magnetic direction.
gpsImgDirection	The direction of the image when it was captured. The range of values is from 0.00 to 359.99.
gpsMapDatum	The geodetic survey data used by the GPS receiver. If the survey data is restricted to Japan, the value of this tag is ‘TOKYO’ or ‘WGS-84’. If a GPS Info tag is recorded, it is strongly recommended that this tag be recorded.
gpsDestLatitudeRef	Reference for latitude of destination
gpsDestLatitude	Latitude of destination, expressed as three values giving the degrees, minutes, and seconds, respectively.
gpsDestLongitudeRef	Reference for longitude of destination
gpsDestLongitude	Longitude of destination, expressed as three values giving the degrees, minutes, and seconds, respectively.
gpsDestBearingRef	Indicates the reference used for giving the bearing to the destination point. ‘T’ denotes true direction and ‘M’ is magnetic direction.
gpsDestBearing	The bearing to the destination point. The range of values is from 0.00 to 359.99.
gpsDestDistanceRef	Indicates the unit used to express the distance to the destination point. ‘K’, ‘M’ and ‘N’ represent kilometers, miles and knots.
gpsDestDistance	The distance to the destination point.
gpsProcessingMethod	A character string recording the name of the method used for location finding. The first byte indicates the character code used, and this is followed by the name of the method.

Element	Description
gpsAreaInformation	A character string recording the name of the GPS area. The first byte indicates the character code used, and this is followed by the name of the GPS area.
gpsDateStamp	date and time information relative to UTC (Coordinated Universal Time). The record format is “YYYY:MM:DD” while converted to W3C-DTF to use in RDF
gpsDifferential	Indicates whether differential correction is applied to the GPS receiver.
interoperabilityIndex	Indicates the identification of the Interoperability rule. ‘R98’ = conforming to R98 file specification of Recommended Exif Interoperability Rules (ExifR98) or to DCF basic file stipulated by Design Rule for Camera File System. ‘THM’ = conforming to DCF thumbnail file stipulated by Design rule for Camera File System.
interoperabilityVersion	Interoperability Version
relatedImageFileFormat	Related image file format
relatedImageWidth	Related image width
relatedImageLength	Related image length
printImageMatching_IFD_Pointer	A pointer to the print image matching IFD
pimContrast	Contrast info for print image matching
pimBrightness	Brightness info for print image matching
pimColorBalance	ColorBalance info for print image matching
pimSaturation	Saturation info for print image matching
pimSharpness	Sharpness info for print image matching

DANUBIUS-PP

Preparatory Phase for the pan-European Research Infrastructure DANUBIUS–RI "The International Centre for advanced studies on river-sea systems"

European Commission

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 739562