

Metadata Ingest and Management (CPP-016)

CPP-Identifier	CPP-016
CPP-Label	Metadata Ingest and Management
Author	Kris Dekeyser
Contributors	Johan Kylander
Evaluators	Felix Burger, Maria Benauer
Date of edition completed	29.08.2025
Change history	Comments
Version 1.0 - 29.08.2025	Milestone version

1. Description of the CPP

The TDA ingests and manages all required *Metadata* including *Metadata* appropriate for specific content types (e.g. geospatial, audio visual).

Inputs and outputs

Input(s)	
Data	<i>File</i> or <i>Object</i>
Metadata	Any <i>Metadata</i> from other process
Documentation / guidance	Metadata recording policy
Output(s)	
Metadata	<i>Provenance metadata</i>
	<i>Technical metadata</i>
	<i>Descriptive metadata</i>
	<i>Structural metadata</i>
	<i>Rights metadata</i>

Definition and scope

Metadata Ingest and Management is about ingesting, storing and maintaining *Metadata* to ensure the long-term preservation of the digital content. This *Metadata* (i.e. “data about the data”) provides contextual information about the *Object* and describes what the *Object* is; how it was created or acquired; how it should be managed and preserved; and how it can be accessed and used in the future. There are different categories of *Metadata* relevant for preservation:

- *Descriptive metadata*: identifies the content, provides contextual information and serves as finding aid;
- *Technical metadata*: contains details on the file formats, features and hard-/software used, compression, etc.;
- *Fixity metadata*: stores information regarding the bitwise state of a *File*. It is used to help to detect any changes made to the *File*’s data;
- *Provenance metadata*: records the origin of the data and keeps track of any processes performed on the data;
- *Structural metadata*: stores the relationships between *Files* and logical parts.
- *Rights metadata*: contains all relevant information about the rights to retain, manipulate and reproduce the data.

The *Metadata* is to be stored in a format that complies with an open and commonly used standard. Most categories of *Metadata* have their own specialised metadata formats and it is therefore not required to store all *Metadata* in a single format. Some common metadata standards are:

- **PREMIS** for *Provenance metadata*
- **METS** for *Structural metadata*
- **Dublin Core, MODS** for *Descriptive metadata*

The *Metadata* can be assigned to *Objects* or *Files* and it must at all times be clear to what *Object* or *File* it is assigned. Some parts of the *Metadata* will be static (e.g. checksums and *File* size) while others will be dynamic and continue to be enhanced or updated during the life cycle of the *Object* in the TDA. It is important for the TDA to keep the *Metadata* safe at all times. No matter how well the data is protected by redundant copies and off-site backups, if *Metadata* is missing the data can no longer be identified and loses its context and thus most of its value.

New *Metadata* is usually created and added during the ingest phase, alongside the *Objects* it relates to. Several processes can and will update *Metadata* during the preservation. If the *Metadata* changes in a way that alters the understanding or interpretation of the *Objects* it preserves, the TDA creates new *AIP* versions that include the updated *Metadata*.

In the *Provenance metadata*, event *Metadata* serves as witness of the execution of the processes performed on the data. A TDA typically documents the preservation actions it performs as *Provenance metadata*, thus serving as an audit trail of the *Objects*. Such event *Metadata* contains:

- An identifier and description of the process;
- A timestamp indicating when the process execution happened;
- Optionally extra *Metadata* associated to the event:
 - Event outcome: Single value data that documents the result of the process. (e.g. 'Success', 'OK', 'Virus free'). The value is typically limited by a controlled vocabulary;
 - Process information: *Metadata* set that documents the tool and environment that was used to perform the process. It will include the tool's name, its version and any configuration parameters that can influence its outcome like virus database version, plugin version, execution parameters, etc.
 - More detailed event result information like ignored warnings, multiple results outcomes, etc.

The *Metadata* can be stored in many ways¹. For example, alongside the data *Files* in a *METS File*, in a relational database as fields, records and tables, in a noSQL database as documents, in a RDF store as a graph of triples or any combination of the above. The storage technology is not relevant as long as the *Metadata* 1) can be searched, retrieved and updated through open and well-documented common standards, and 2) is clearly and consistently linked to the *Object* or *File*. Usage of identifiers and URIs are key.

¹ See also PREMIS v3 - pg 25 "Storing metadata"

Process description

Trigger event(s)

Trigger event	CPP-identifier
Any process that generates event timestamps, optionally including outcome data	Most CPPs
A new checksum was generated	CPP-001 Checksum Generation and Validation
A checksum was validated	CPP-003 Integrity Checking
A new identifier was generated and attached to an <i>Object</i>	CPP-005 Identifier Management
A format was assigned to a <i>File</i>	CPP-008 File Format Identification
<i>Metadata</i> was extracted from a <i>File</i>	CPP-009 Metadata extraction
Missing <i>Metadata</i> has been identified, or used metadata standards have become obsolete	CPP-012 Risk Mitigation
<i>Object</i> is removed	CPP-017 Disposal
<i>Object</i> rights changed	CPP-020 Rights management
<i>Descriptive metadata</i> must be pseudonymised/anonymised	CPP-020 Rights management

Step-by-step description

No	Supplier	Input	Steps	Output	Customer
----	----------	-------	-------	--------	----------

1		Event with optional data	Store event occurrence with timestamp and extra data	<i>Provenance metadata</i>	
2	CPP-001 Checksum Generation and Validation	New checksum(s) created	Store checksums with their respective algorithms and assign it to the given <i>File</i>	<i>Fixity metadata</i>	
3	CPP-003 Integrity Checking	Checksum was validated	Update the checksum's last validation timestamp	<i>Fixity metadata</i>	
4	CPP-005 Identifier Management	New identifier	Store the identifier and assign to the <i>Object</i> or <i>File</i>	<i>Descriptive metadata</i>	
5	CPP-008 File Format Identification	Format identifier and format registry identifier	Store the format information and assign to the <i>File</i>	<i>Technical metadata</i>	
6	CPP-009 Metadata extraction	Any <i>Metadata</i> extracted from the <i>File</i>	Store the <i>Metadata</i> and assign to the <i>File</i>	<i>Technical metadata</i>	
				<i>Provenance metadata</i>	
				<i>Descriptive metadata</i>	
				<i>Structural metadata</i>	
				<i>Rights metadata</i>	
7	CPP-017 Disposal	Set of <i>Metadata</i>	Replace the <i>Object</i> or <i>File Metadata</i> with the minimal set and remove all references to the <i>Object</i> and <i>Files</i>	Any <i>Metadata</i>	

8	CPP-020 Rights management	New set of rights information	Store the rights information and assign to the <i>Object</i> or <i>File</i>	<i>Rights metadata</i>	
9	CPP-020 Rights management	Requirement to anonymise/pseudonymise personal data	Update <i>Descriptive metadata</i> and create new <i>AIP</i> version	<i>Descriptive metadata</i>	
10	CPP-012 Risk Mitigation	Set of <i>Metadata</i>	Used metadata standard has become obsolete and must be migrated	<i>Any Metadata</i>	
11		Set of <i>Metadata Information Packages</i>	If needed (after updating the <i>Metadata</i>): The TDA creates a new version of an <i>AIP</i> to store the new or updated <i>Metadata</i>	<i>AIP</i> version	CPP-021 (AIP versioning)

Rationale(s)² and worst case(s)

Rationale	Impact of inaction or failure of the process
FAIR F3 Metadata clearly and explicitly include the identifier of the data they describe	If the connection between <i>Metadata</i> and the <i>Object's</i> data is lost, <i>Objects</i> can no longer be found by searching the <i>Metadata</i> .
FAIR F4 (Meta)data are registered or indexed in a searchable resource	Without searchability it is not possible to retrieve sets of relevant data based on characteristics described in the <i>Metadata</i> .
FAIR A2 Metadata should be accessible even when the data is no longer available	Storing a tombstone metadata for disposed content serves as a witness of the data's existence beyond its disposal.
FAIR I1 (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation FAIR R1.3 (Meta)data meet domain-relevant community standards	Lack of understanding the language in which the <i>Metadata</i> is expressed leads to misinterpretation or even loss of information.

2. Dependencies and relationships with other CPPs

Dependencies

CPP-ID	CPP-Title	Relationship description
Many CPPs		Many CPPs generate events and related data that needs to be stored safely as <i>Provenance metadata</i> and linked to the <i>Object</i> or <i>File</i> .
CPP-001	Checksum Generation and Validation	The checksums and associated algorithms need to be stored in the <i>File's Fixity metadata</i> .
CPP-003	Integrity Checking	The timestamp of the <i>File's</i> checksum needs to be updated to keep track of the last successful check.
CPP-005	Identifier	The TDA must store the <i>Persistent Identifier</i> and link it to

² Term derived from PREMIS.

	Management	the <i>Object</i> or <i>File</i> .
CPP-008	File Format Identification	Format information needs to be stored with the <i>File</i> 's <i>Technical metadata</i> .
CPP-009	Metadata extraction	Any <i>Metadata</i> that was extracted from the <i>File</i> needs to be stored, searchable and retrievable.
CPP-017	Disposal	A tombstone consisting of <i>Metadata</i> only - without the actual data - has to be kept as a witness of the data's former presence.
CPP-020	Rights management	Any information regarding the rights for the TDA and the end users on the data needs to be stored and should be searchable and retrievable.

Other relations

Relation	CPP-ID	CPP-Title	Relationship description
Required by	CPP-013	Object Management Reporting	<i>Metadata</i> ingest provides new <i>Provenance metadata</i> .
Required by	CPP-024	Enabling Discovery	Enabling Discovery relies on a correct <i>Metadata</i> management process. In particular, <i>Metadata</i> created by and within the TDA is of particular interest to the consumer in order to understand preservation actions that could have affected the <i>Object</i> .
Required by	CPP-029	Ingest	The ingest process produces <i>Technical</i> , <i>Rights</i> and <i>Provenance metadata</i> that are recorded in the <i>Information package</i> and digital archive database by Metadata Ingest and Management.
Affinity with	CPP-022	Significant properties definition	A TDA defines significant properties for digital <i>Objects</i> . These are then translated to <i>Technical metadata</i> that is ingested. The significant properties definition process also influences which technical metadata standards are applied.

3. Links to frameworks

Certification

Certification framework	Term used in framework to refer to the CPP	Section
CTS Link	Provenance and authenticity Storage & Integrity	R07 R14
Nestor Seal Link	Logging the preservation measures	C31, page 49
ISO 16363 Link	Information Management	Section 4.5, page 4-26

Other frameworks and reference documents

Reference Document	Term used in framework to refer to the process	Section
OAIS Link	/	OAIS mentions <i>Metadata</i> , but does not discuss metadata ingest or metadata management
PREMIS Link	Storing metadata	Section with the same title in the introduction

4. Reference implementations

Publicly available documentation

Institution	Organisation type	Language	Hyperlink
TIB – Leibniz Information Centre for Science and Technology and University Library, Germany	National library	English	https://wiki.tib.eu/confluence/spaces/lza/pages/93608951/Metadata
	Non-commercial digital preservation service		
	Research infrastructure		
	Research performing organisation		
CSC – IT Center for Science Ltd., Finland	Non-commercial digital preservation service	Finnish	https://urn.fi/urn:nbn:fi-fe2024051731943 (Annex 4, section 2.2.1)
		English	https://urn.fi/urn:nbn:fi-fe2020100578094 (section 2.4.4. and 3.3.)
Archivematica	Digital preservation system	English	https://www.archivematica.org/en/docs/archivematica-1.17/user-manual/transfer/transfer/#transfers-with-metadata ; (Import metadata)