

# AIP Versioning (CPP-021)

<b>CPP-Identifier</b>	CPP-021
<b>CPP-Label</b>	AIP Versioning
<b>Author</b>	Bertrand Caron
<b>Contributors</b>	Micky Lindlar, Kris Dekeyser
<b>Evaluators</b>	Matthew Addis, Maria Benauer, Fen Zhang
<b>Date of edition completed</b>	29.08.2025
<b>Change history</b>	<b>Comments</b>
Version 1.0 - 29.08.2025	Milestone version

# 1. Description of the CPP

The TDA creates successive versions of the *Information Objects* on its own or on the producer initiative.

## Inputs and outputs

Input(s)		
Data	<i>AIP</i>	
	Optional: <i>SIP</i> (if Versioning is triggered by an updating request the Producer)	
	Optional: <i>File(s)</i> or <i>Representation(s)</i> that should replace existing ones or be added to an <i>AIP</i>	
Metadata	<i>Metadata</i> that should be added to the <i>Information Package</i> (i.e. <i>Descriptive</i> , <i>Technical</i> , <i>Structural</i> and/or <i>Administrative metadata</i> )	
Documentation / guidance	Versioning Policy	
	<i>AIP</i> packaging policy	
Output(s)		
Data	New <i>AIP</i> Version	
Metadata	<i>Provenance metadata</i>	
	Storage management information (new version identifier)	
	Updated <i>Metadata</i>	<i>Descriptive metadata</i>
		<i>Technical metadata</i>
		<i>Structural metadata</i>

## Definition and scope

Updating *AIPs* is a common operation in a TDA that involves keeping track of changes to data or *Metadata*, and preserving previous versions (entirely or partially) in order to go back to different stages of an *AIP*. Either the producer or the TDA may initiate an update, usually because of technical or contextual changes on the level of *File*, *Representation* or *Metadata*.

Reasons for Versioning include, but are not limited to:

For contextual quality reasons:

- Removal of data
  - Fault in the initial deposit: An *AIP* or *Representation* contains *File(s)* that were not supposed to be part of the deposit (e.g. system *Files* not needed for preservation);
  - Content being retracted after the initial deposit: An *AIP* or *Representation* contains *File(s)* that were withdrawn (e.g. due to security reasons, legal restrictions).
- Update of data
  - Update of *File(s)*: An *AIP* or *Representation* contains *File(s)* that have been replaced (e.g. due to faults in the initial deposit);
  - Enrichment or correction of *Descriptive Metadata*;
  - Update of administrative *Metadata* (e.g. due to changes in the access rights or legal information for an *AIP*).
- Addition of data
  - Incomplete initial deposit: The initial *Information Package* was missing *Files* and some need to be added after its ingest;
  - Enhancement of the initial deposit: After ingest of the *Information Package (IP)*, additional material was issued and needs to be preserved in the *IP*.

For technical quality reasons:

- Replacement or Update due to preservation action: A *File* or *Representation* has been migrated into a format deemed more suitable for preservation;
- Replacement or Update due to new packaging: A *File* or *Representation* has been re- or unpackaged (e.g. for ZIP, TAR, WARC or disk image containers).

For *Metadata* quality reasons:

- Addition or replacement of *Technical metadata* as a consequence of re-running a characterisation process (i.e. Format Identification, Metadata Extraction or Format Validation) on already ingested *AIPs* (e.g. due to a new tool being released).

Updates might involve the submission of a new *SIP* by the producer (e.g. in the case of changes to the initial deposit), but can also just take place on the storage level (e.g. new data being added as a result of a preservation action) without a *SIP* being involved.

In comparison to other IT domains, updating and creating new versions in digital preservation exhibits peculiarities for the following reasons:

- On the *File* level: As data and *Metadata* are generally stored in physical containers (such as ZIP, TAR, WARC, etc.), unpacking and merging content from the incoming *SIP* is a complex and risky process;
- On the storage level: As physical containers corresponding to *IPs* are usually stored on tape or other cold-storage technologies, updates are generally asynchronous;
- On the *Metadata* level: Authenticity and audit trail requirements imply that TDAs record precisely all details about the versioning impacts (e.g. requester, date, reasons, impacted data and *Metadata*, etc.). This is of particular importance when the versioning is requested by TDA and when the new version replaces the original one.

A TDA must have a versioning policy in place for handling and documenting retracted or updated data. It may decide to partially or totally retain the older versions, based on several factors such as reversibility (e.g. addition of new *Files* versus replacement of *Files*), purpose of the update (e.g. correction, enrichment, risk mitigation, etc.), issuer of the updating request (i.e. Producer or TDA), etc.

The implementation of updates to *AIP* (i.e. data and/or *Metadata*) requires consideration of a range of factors including a) what is being updated and with what frequency, b) whether that should result in a new *AIP* version or simply an update to the current version, c) how often

the different versions of the *AIP* need to be accessed, d) how the *AIP* is stored internally (e.g. using databases, file systems, tape libraries, cloud storage), and e) how versions are serialised and exported from the TDA when they need to be exchanged with third-parties.

Specific implementation approaches are outside of the scope of this CPP.. However, some considerations may include:

- Are there regular and incremental updates in dynamic collections of content? If so, consider storing the changes (deltas) between *AIP* versions, rather than complete *AIP*s, which can reduce both storage requirements and processing overhead.
- Are there any changes to the *AIP* content that do not require versioning (e.g. recording the results of fixity checks). If so, consider mechanisms that allow efficient tracking of preservation actions, such as preservation *Metadata* updates, without duplicating or accessing unchanged content.
- Is there a need for enhanced traceability and provenance so that it is easy to follow the evolution of *AIP*s over time, without needing to access or analyse internal *Metadata*? If so, consider additional indexing and logging mechanisms alongside *AIP* storage.
- The ability to access the deltas between *AIP*s instead of whole versions can support easier integrity checks and rollback capabilities (e.g. enabling recovery from data corruption without full *AIP* restores).
- Is there a need to externalise and exchange *AIP*s with third parties (e.g. repository migration, synchronisation, and replication)? Exchanging *AIP* versions as deltas can enable more efficient data transfer across distributed or replicated preservation systems.
- If some or all of *AIP*s are held using offline media, then minimising data handling can be desirable to reduce the risk of corruption or damage when media is retrieved or read. Isolating and preserving only changed components of *AIP*s can reduce unnecessary manipulation of stable data, which lowers the risk of inadvertent corruption, loss, or operational error.

## Process description

### Trigger event(s)

Trigger event	CPP-identifier
Updating request from the Producer in the form of an <i>SIP</i> ingest request	
Data must be removed or redacted because of legal constraints	CPP-020 (Rights Management)
Re-run of a characterisation process (e.g. when a new tool or new tool version is made available)	CPP-008 (File Format Identification), CPP-009 (Metadata Extraction), CPP-010 (File Format Validation)
New <i>Representations</i> created by the TDA need to be added or should replace the original	CPP-014 (File Migration), CPP-027 (File Repair), CPP-028 (Creation of Derivatives)

### Step-by-step description

No	Supplier	Input	Steps	Output	Customer
1			Retrieve corresponding <i>AIP Metadata</i>		
2			Identify the scope of versioning	Update of <i>Metadata</i> (step 3a)	
				Addition or removal of <i>Files</i> (step 3b)	
				Replacement of an <i>AIP</i>	

				(step 3c)	
3a		Additional incoming <i>Metadata</i>	If the intended update concerns only <i>Metadata</i> , update the <i>AIP Metadata</i>	Digital archive database update	
		<i>AIP Metadata</i>			
3b		<i>SIP</i> (if relevant)	If the update involves the addition of <i>Files</i> , request corresponding physical <i>AIP</i> and perform merging with incoming data and <i>Metadata</i>	Content of the new <i>AIP</i> version	
		<i>AIP</i>		Digital archive database update	
		<i>File(s)</i> or <i>Representation(s)</i> that should replace existing <i>File(s)</i> or be added to the <i>AIP</i>			
		Versioning policy			
3c		Versioning policy	If the update is a “full update”, ingest data and <i>Metadata</i> as a complete replacement of the previous <i>AIP</i> version		
4	CPP-005 (Identifier Management)	New PID for the new <i>AIP</i> version	Request new identifier for the new <i>AIP</i> version	<i>AIP</i> version with PID assigned	
5			Document the date, requester, data or <i>Metadata</i> impacted, and reason for the versioning	<i>Provenance metadata</i>	

6		Content of the new <i>A/P</i> version	Repackage the content of the new <i>A/P</i> according to the TDA's current packaging standards	New <i>A/P</i>	CPP-029 (Ingest)
		<i>A/P</i> packaging policy			
7	CPP-012 (Risk Mitigation)	Versioning policy	Manage the retention of previous version(s) (i.e. partial, total retainment or disposal)		

## Rationale(s)<sup>1</sup> and worst case(s)

Rationale	Impact of inaction or failure of the process
Digital <i>Objects</i> are dynamic and easy to modify, correct or enrich. Producers generally use this particularity to update them for various reasons. A TDA should be able to manage versioning in order to identify and preserve successive <i>AIP</i> versions.	Inability to handle updating requests from the Producer causes duplication because the TDA needs to store subsequent <i>AIP</i> versions as new <i>AIP</i> s. See Caron, B. and Verrier, L. 2024. From Products to Library Collections: Towards a Data-driven Policy for Legal Deposit of Born-digital Sound at the National Library of France. iPRES 2024 Papers - International Conference on Digital Preservation. Available at <a href="https://ipres2024.pubpub.org/pub/6m5xlcii/release/1">https://ipres2024.pubpub.org/pub/6m5xlcii/release/1</a> .
Producers may want to perform complex updating operations involving several changes to different parts of an <i>AIP</i> and combine all changes into a single version.	Successive, incomplete changes performed by the Producer may lead to the TDA's inability to identify stable and complete versions.

## 2. Dependencies and relationships with other CPPs

### Dependencies

CPP-ID	CPP-Title	Relationship description
CPP-005	Identifier Management	Every new <i>AIP</i> version must be assigned a new identifier. The design and versioning of identifiers (e.g. creation of entirely new identifiers, use of version qualifiers etc.) should be defined in a persistent identifier minting policy.
CPP-009	Metadata Extraction	Soft dependency (i.e. may require): The documented event, datetime, and <i>Provenance metadata</i> from the Metadata Extraction process may be required by <i>AIP</i> Versioning.
CPP-012	Risk Mitigation	Risk Mitigation acts as a supplier to <i>AIP</i> Versioning, providing <i>AIP</i> Versioning with risk mitigation policy details, as they relate to managing risks in retention of previous version(s) (i.e. partial, total retainment or disposal).

---

<sup>1</sup> Term derived from PREMIS.



CPP-029	Ingest	Versioning implies several delicate operations, in particular in the case of a partial update, where the incoming <i>SIP</i> should be merged with the existing <i>AIP</i> .
---------	--------	--

## Other relations

Relation	CPP-ID	CPP-Title	Relationship description
Affects	CPP-006	AIP Batch Export	Versioning impacts how the export will have to be run and where and how information about the versions may be found. In addition a policy might determine if only the last or all versions should be exported.

## 3. Links to frameworks

### Certification

Certification framework	Term used in framework to refer to the CPP	Section
CTS <a href="#">Link</a>	versions	Provenance and authenticity (R07) The repository guarantees the authenticity of the digital objects and provides provenance information.” includes “The repository approach to changing and versioning data and metadata. How the approach and records of changes are communicated to data depositors and users.“
Nestor Seal <a href="#">Link</a>	versions	Topic not addressed, but briefly mentioned in C27 Identification
ISO 16363 <a href="#">Link</a>	versions	AIP Versions and Editions are essential notions of the OAIS / ISO 16363 model. No AIP versioning operation is described, but the notion of versions is touched upon in the following sections: <ul style="list-style-type: none"> <li>- 4.2.1.1 The repository shall have a definition of each AIP or class of AIPs that is adequate for long-term preservation, enabling the identification, extraction and interpretation of all the required</li> </ul>

		<p>components within each AIP.</p> <ul style="list-style-type: none"> <li>- 4.3.5 The repository must be able to export the relevant AIPs to its successor location in the timescale required, together with supplementary information including specifications of the AIPs adequate for the information contained in them to be extracted.</li> <li>- 4.4.1.1 The repository shall preserve the Content Information of AIPs</li> <li>- 5.1.2 The repository shall manage the number, security, coordination, and location of copies of all digital objects.</li> </ul>
--	--	---

## Other frameworks and reference documents

Reference Document	Term used in framework to refer to the process	Section
OAIS	AIP Version AIP Edition	<p>Section 5.2.4.6 “Distinguishing AIP Versions, AIP Editions and Derived AIPs”</p> <p>OAIS distinguishes AIP Versions (triggered by migrations performed by the Archive) and AIP Editions (triggered by update requests from the Producer).</p>
PREMIS	/	<p>PREMIS does not approach versioning. Nevertheless, the “1:1 principle” implies that if an Object (Representation, File or Bitstream) is modified by the repository, this operation creates a new Object that is declared as being derived from the source Object.</p>

## 4. Reference implementations

### Example use case(s)

#### BnF Information Package lifecycle Implementation

Institutional Background	
Institution	Bibliothèque nationale de France, France
Hyperlink	Bertrand Caron, Jordan de La Houssaye, Thomas Ledoux, Stéphane Reecht. Life and Death of an Information Package: Implementing the Lifecycle in a Multi-Purpose Preservation System. <i>iPRES 2017 14th International Conference on Digital Preservation</i> , Sep 2017, Kyoto, Japan. <a href="#">hal-01617645</a>
Description	
Trigger event	The paper describes several trigger events (update, deletion or simple <i>Metadata</i> edition requests) and different scenarios for versioning.
Problem statement	The paper describes scenarios for partial updates which may have varying impacts on existing <i>AIPs</i>
Proposed solution	Following an unsuccessful attempt to implement an algorithm to determine the nature of update operations, based on structural comparisons between incoming <i>SIPs</i> and existing <i>AIPs</i> , BnF adopted a policy that permits only full updates. As a result, producers are required to request the relevant <i>AIP</i> and take responsibility for merging it with the <i>SIP</i> according to their needs.

## Publicly available documentation

Institution	Organisation type	Language	Hyperlink
TIB – Leibniz Information Centre for Science and Technology and University Library, Germany	National library	English	<a href="https://knowledge.exlibrisgroup.com/Rosetta/Product_Documentation/Rosetta_Staff_Users_Guide/Web_Editor/001_Web_Editor/003_Viewing_Digital_Objects_in_the_Web_Editor#Versions_Tab">https://knowledge.exlibrisgroup.com/Rosetta/Product_Documentation/Rosetta_Staff_Users_Guide/Web_Editor/001_Web_Editor/003_Viewing_Digital_Objects_in_the_Web_Editor#Versions_Tab</a> and <a href="https://wiki.tib.eu/confluence/spaces/lza/pages/93608951/Metadata#Metadata-EventLoggingofpreservationactions">https://wiki.tib.eu/confluence/spaces/lza/pages/93608951/Metadata#Metadata-EventLoggingofpreservationactions</a>
	Non-commercial digital preservation service		
	Research infrastructure		
	Research performing organisation		
Archivematica	Digital preservation system	English	<a href="https://www.archivematica.org/en/docs/archivematica-1.17/user-manual/archival-storage/aic/#aic">https://www.archivematica.org/en/docs/archivematica-1.17/user-manual/archival-storage/aic/#aic</a> ; and <a href="https://www.archivematica.org/en/docs/archivematica-1.17/user-manual/ingest/ingest/#reingest">https://www.archivematica.org/en/docs/archivematica-1.17/user-manual/ingest/ingest/#reingest</a> ; and <a href="https://www.archivematica.org/en/docs/archivematica-1.17/user-manual/preservation/preservation-planning/#">https://www.archivematica.org/en/docs/archivematica-1.17/user-manual/preservation/preservation-planning/#</a>