

# Creation of Derivatives (CPP-028)

<b>CPP-Identifier</b>	CPP-028
<b>CPP-Label</b>	Creation of Derivatives
<b>Author</b>	Kris Dekeyser
<b>Contributors</b>	Juha Lehtonen
<b>Evaluators</b>	Matthew Addis, Felix Burger
<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 generates derivative copies to address the specific needs of its *Designated Community*.

## Inputs and outputs

Input(s)	
Data	Source <i>Representation</i> of an <i>Object</i>
Documentation / guidance	File format policy - Derivatives
Output(s)	
Data	A new <i>Representation</i> , a new <i>AIP</i> , or a <i>DIP</i>
Metadata	<i>Provenance metadata</i>
	<i>Structural metadata</i>

## Definition and scope

The process of Creation of Derivatives refers to generating new *Representations* of a digital *Object* (based on the original *Representation* of the *Object*) to address specific needs of the Designated Community. Some of these needs are:

- Provide easy access to the information for a broader community (e.g. conversion from TIFF to JPEG in order to allow viewing in a web browser)
- Provide access for a specific user base (e.g. converting CAD *Files* to a specific format that is more commonly in use by a group of researchers)
- Provide different forms of access to the data (e.g. combining a tree of HTML-pages, images and other resources into a WARC *File*; unpacking a ZIP *File*; compiling data into a pivot table or chart)
- Provide access to the data in formats that address accessibility challenges (e.g. creating DAISY derivatives of PDFs for visually impaired consumers)
- Restricting or protecting the access to the data (e.g. creating a PDF from an e-book containing only the first few pages; lowering the resolution and watermarking; thumbnails)
- Enriching the information (e.g. adding AI-generated transcriptions as subtitles to a video; overlaying transparent OCR text over a scanned image in a PDF; adding AI-generated *Descriptive metadata*)

Creation of Derivatives differs from other processes that create new *Files* or *Representations*:

- Unlike **File Migration** (CPP-014), Creation of Derivatives reproduces only the information and significant properties of the original that is useful to address specific needs of the Designated Community. Thus, the output of Creation of Derivatives may lack part of the information that is not required to satisfy said needs.
- **File Normalisation** (CPP-026) is performed during ingest and aims at reducing the number of formats by converting *Files* or *Representations* to preservation formats defined by the institutional formats policy.

- **File Migration** (CPP-014) is performed after ingest to address a specific preservation risk.

Derivative copies can be short-lived and there can be more than one derivative copy for a given original. For example, derivative copies for a video master *File* could include multiple resolutions, different levels of compression and different encoding schemes (e.g. suited for previewing rather than reuse). The derivative videos may be generated to support contemporary web delivery standards such as HTML5 and as a result may be short lived compared to the copies created as a result of migration and/or normalisation.

It is up to the TDA to decide how far they take this and if they support any of these use cases, if at all. The generation of the derivative copies can be performed during delivery, during or before the ingestion or any time during the *Object's* life cycle in the TDA. Performance, *Object* size, availability of conversion tools and workflows, and user requirements will be some of the elements that could play a role in this decision.

Depending on whether the derivative copy maintains the significant properties of the original *Object*, the derivative copy may then (a) be added to the same *AIP* as the original (e.g. as a new *Representation*), (b) result in a new *AIP* that is stored in addition to the original and is linked to it, or (c) may be held in a *DIP* which in turn may be stored temporarily or permanently, and is linked to the original *AIP*. In PREMIS terms, the derivative copies can be stored as a new *Representation* if the derivative copy is a complete and meaningful rendition of the *Intellectual Entity*. However, if the derivative copies change the meaning and of the original (e.g. significant properties are not maintained) then a new *AIP* or *DIP* is recommended.

## Process description

### Trigger event(s)

Trigger event	CPP-identifier
Request for access to a digital <i>Object</i>	CPP-025 (Enabling Access)
<i>SIP</i> processing during Ingest	CPP-029 (Ingest)

### Step-by-step description

No	Supplier	Input	Steps	Output	Customer
1A	CPP-025 Enabling Access	Request for access copy and requested <i>Object(s)</i>	Identify the <i>Representation</i> to be processed	<i>Representation</i>	
1B	CPP-029 Ingest	Ingest of a <i>SIP</i>	Iterate over the <i>Representations</i> in the <i>SIP</i>	<i>Representation</i>	
2		<i>Representation</i>	Determine if a derivative is needed by applying the rules in the policy to the <i>Files</i> in the <i>Representation</i> . End the process unless a derivative is needed		
		File format policy - Derivatives			
3		<i>Representation</i>	Determine the migration path and apply it to the source <i>File(s)</i> in the <i>Representation</i>	New <i>File</i> or set of <i>Files</i>	
		File format policy - Derivatives: tools			

4		File format policy - Derivatives	Decide on the packaging of the <i>File(s)</i> : a new <i>Representation</i> , a new <i>AIP</i> or a <i>DIP</i> .	<i>Representation, AIP or DIP</i>	
5A		<i>DIP</i>	Deliver the <i>DIP</i> and end the process		
5B	CPP-001 (Checksum Generation and Recording) CPP-008 (File Format Identification) CPP-009 (Metadata Extraction) CPP-010 File Format Validation)	<i>Representation or AIP</i>	Apply the characterisation processes against the new <i>Object</i> , optionally by running the ingest process and all its subprocesses	Validated <i>Object</i>	CPP-015 (Identifier Management)
6			Report migration performed, success/failure, <i>Files</i> generated	Audit trail Provenance metadata Structural metadata	

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

Rationale	Impact of inaction or failure of the process
<p>FAIR Interoperable: I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.</p> <p>FAIR Reusable: R1.3. (Meta)data meet domain-relevant community standards</p> <p>i.o.w. To support Interoperability and Reuse of the data, it should be made available in a format or structure that closely matches its usage.</p>	<p>The consumer of the data may resort to other tools to convert it and corrupt or distort the data in the process.</p> <p>The consumer of the data may be unable to use the data because it is not in a format they can use and they do not have the tools/knowledge to convert it themselves.</p>
FAIR Findable	Without derivative copies, the consumer of the data may find it difficult and time consuming to assess whether the data contains the content they need, for example they cannot easily preview/inspect derivatives that allow them to determine if they should go ahead and retrieve the full dataset.

## 2. Dependencies and relationships with other CPPs

### Dependencies

CPP-ID	CPP-Title	Relationship description
CPP-022	Significant Properties Definition	Though derivative copies may not be subject to the same requirements as preservation copies, providing access through copies that do not convey significant properties might be considered a TDA failure.
CPP-025	Enabling Access	The request for access can trigger the creation of a derivative copy for rendering purposes.
CPP-005	Identifier Management	Soft dependency (i.e. may require): A derivative copy of a <i>File</i> may get its own identifier.
CPP-029	Ingest	Soft dependency (i.e. may require): During ingest derivative copies may be created.

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

## Other relations

Relation	CPP-ID	CPP-Title	Relationship description
Required by	CPP-013	Object Management Reporting	Creation of new <i>Objects</i> provides data for statistical reporting.
Facilitates	CPP-015	Emulation and Rendering Tools	Derived copies in a different format or structure than the original <i>Representation</i> can help in rendering the data or simulating the environment.
Not to be confused with	CPP-014	File Migration	Creation of a derivatives creates an additional <i>Representation</i> to be retained, while the migration outcome is supposed to replace its source.
Not to be confused with	CPP-026	File Normalisation	Unlike File Normalisation (CPP-026), Creation of Derivatives (CPP-028) reproduces only the information and significant properties of the original that is useful to address specific needs of the <i>Designated Community</i> . Thus, the output of Creation of Derivatives may lack part of the information that is not required to satisfy said needs. Normalisation aims to maintain all the significant properties of the original.

## 3. Links to frameworks

### Certification

Certification framework	Term used in framework to refer to the CPP	Section
CTS <a href="#">Link</a>	/	/
Nestor Seal <a href="#">Link</a>	/	/
ISO 16363 <a href="#">Link</a>	dissemination of digital objects	4.6.2 dissemination may simply be a copy, a simple format transformation or a more complex derivation

## Other frameworks and reference documents

Reference Document	Term used in framework to refer to the process	Section
OAIS <a href="#">Link</a>	<i>DIP</i> and derived <i>AIP</i>	OAIS differentiates between generating derivations for the sake of a <i>DIP</i> and creating derivative copies and storing them in the TDA. In the latter case, they suggest it is added as an API linked to the source <i>AIP</i> , ignoring the fact that it could be stored as a <i>Representation</i> within the same <i>AIP</i> .
PREMIS <a href="#">Link</a>	Relationships	Derivation relationships can be defined between <i>Objects</i>



## 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	part of pre-ingest routines, see: <a href="https://wiki.tib.eu/confluence/spaces/lza/pages/93608628/Pre-Ingest?preview=/93608628/93608629/pre-ingest.png">https://wiki.tib.eu/confluence/spaces/lza/pages/93608628/Pre-Ingest?preview=/93608628/93608629/pre-ingest.png</a> and <a href="https://wiki.tib.eu/confluence/spaces/lza/pages/93608628/Pre-Ingest?preview=/93608628/310285401/film_digitization.png">https://wiki.tib.eu/confluence/spaces/lza/pages/93608628/Pre-Ingest?preview=/93608628/310285401/film_digitization.png</a>
	Non-commercial digital preservation service		
	Research infrastructure		
	Research performing organisation		
Archivematica	Digital preservation system	English	In Archivematica, creating derivatives (called access copies) are part of normalisation. <a href="https://www.archivematica.org/en/docs/archivematica-1.17/user-manual/ingest/ingest/#normalize">https://www.archivematica.org/en/docs/archivematica-1.17/user-manual/ingest/ingest/#normalize</a> ; <a href="https://www.archivematica.org/en/docs/archivematica-1.17/user-manual/preservation/preservation-planning/#normalization">https://www.archivematica.org/en/docs/archivematica-1.17/user-manual/preservation/preservation-planning/#normalization</a>