

Replication (CPP-011)

CPP-Identifier	CPP-011
CPP-Label	Replication
Author	Johan Kylander
Contributors	Bertrand Caron, Juha Lehtonen
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 automatically manages the replication of *Information Packages* to multiple storage locations (potentially in different geographical locations).

Inputs and outputs

Input(s)	
Data	<i>AIP</i>
Metadata	<i>Storage management information</i>
Documentation / guidance	Storage management policy - Copies
Output(s)	
Data	<i>AIP</i>
Metadata	<i>Storage management information</i>

Definition and scope

Replication is the process of copying data to a new storage medium in order to create new identical copies of *AIP(s)*. All new copies created must have their fixity verified to ensure that the process was completed successfully. The process creates new parallel copies to achieve redundancy in the storage system. The process of replication makes no changes to the *AIPs* themselves but only operates on the storage layer to add storage locations for the new copies.

Redundancy is required to mitigate risks such as data corruption and unintentional or unwarranted data destruction by creating a fault tolerance. The number of parallel copies and the choice of media types is determined by how a TDA implements **Risk Mitigation** (CPP-012). Typically, strategies for the following aspects of redundancy and replication should be covered in the TDA policy:

- *Creation of fault tolerance*: The creation of at least three parallel copies is recommended. In this way, the data is always stored on two other valid copies in cases where one copy is corrupted or destroyed;
- *Use of diverse storage media*: Employing different storage media helps to avoid systematic errors and vulnerabilities and reduces the risk of vendor lock-in in a single storage solution;
- *Geographical distribution of copies*: Storing the copies geographically dispersed reduces location based risks such as environmental disasters or political instability;
- *Offline storage of copies*: Storing some copies offline reduces the risk of unwarranted tampering of the data by malicious users or software.

Process description

Trigger event(s)

Trigger event	CPP-identifier
Ingest. New ingested data that needs to be replicated when accepted to preservation.	CPP-029 Ingest
Updates to a TDAs decisions on copy management as per its Risk Mitigation policies. The TDA defines how many copies are needed at all times and on what type of media. Changes to these policies can occur during preservation	CPP-012 Risk Mitigation

Step-by-step description

No	Supplier	Input	Steps	Output	Customer
1	CPP-012 (Risk Mitigation)	Copy management policy	Identify and locate <i>AIP</i> to be replicated	Inventory of <i>AIP</i> to be replicated	
	CPP-029 (Ingest)	Storage management information		Source storage medium with <i>AIP</i>	
		<i>AIP</i>			
2		Copy management policy	Select target storage medium to copy the <i>AIP</i> to	Target storage medium	

3		Source storage medium with <i>AIP</i>	Start the copy process (steps 4 to 8).		
		Target storage medium			
		<i>AIP</i>			
4			Retrieve the <i>AIP</i> from the source storage medium		
5			Copy the <i>AIP</i> to the new storage medium	New copy of <i>AIP</i>	
6		Existing/previous <i>Fixity metadata</i>	Validate the fixity of the <i>AIP</i> on the target storage medium	Valid status (step 7)	<i>Fixity Metadata</i>
				Invalid status (go back to step 4)	
7			Update the fixity for the new <i>AIP</i> copy	<i>Fixity Metadata</i>	
8			Update the storage location for the new <i>AIP</i> copy	<i>Storage management information</i>	
9			Create preservation <i>Event metadata</i> of the replication	<i>Provenance metadata</i>	

Rationale(s)¹ and worst case(s)

Rationale	Impact of inaction or failure of the process
Have multiple copies of data	Handling data corruption or destruction requires intact copies to restore lost data from.
Copies in different storage locations with different location based risks	If data is not stored in different locations, any disaster to the location is a single point of failure. The same applies to the type of risk associated with the locations (e.g. political risk, risk of natural disasters etc.).
Copies should exist on different type of storage media	Systematic errors in a storage media can affect all copies on the same type of media.
Offline copies that are not accessible by normal procedures	Offline copies are safe from malicious users or software, as they cannot be accessed by normal means.
Open source storage solutions	Vendor lock-in in storage solutions can pose a risk to the copies on those media if they have a dependency to an outside partner.

2. Dependencies and relationships with other CPPs

Dependencies

CPP-ID	CPP-Title	Relationship description
CPP-012	Risk Mitigation	A TDAs storage policy, that defines how data is stored, the amount of parallel copies etc, is based on a TDAs risk assessment and mitigation.
CPP-005	Identifier Management	Soft dependency (i.e. may require): When a <i>Digital Object</i> or <i>File</i> is replicated, the replicant may be assigned a new PID.

¹ Term derived from PREMIS.

Other relations

Relation	CPP-ID	CPP-Title	Relationship description
Not to be confused with	CPP-030	Refreshment	Replication creates new parallel copies of <i>AIPs</i> , while Refreshment replaces <i>AIPs</i> onto new storage media.
Affinity with	CPP-002	Checksum Validation	All new <i>AIP</i> copies must have their checksum validated to verify that the process was successful. The checksum validation is more mechanical in its nature in Replication, only aiming at verification of the copy process. In contrast to CPP-002, it does not have to negotiate with producers or examine the results.
Affinity with	CPP-004	Data Corruption Management	A replication-like process is performed in data corruption management, in which <i>AIPs</i> must be replicated to replace corrupted <i>AIPs</i> .
Affinity with	CPP-006	AIP Batch Export	Replication creates new parallel copies of <i>AIPs</i> within a TDAs archival storage. AIP Batch Export exports <i>AIPs</i> to external locations.

3. Links to frameworks

Certification

Certification framework	Term used in framework to refer to the CPP	Section
CTS Link	“multiple copies”	R14 Storage & Integrity
Nestor Seal Link	“selection of suitable storage media, redundancy, refreshing, media migration”	C15 Integrity: Functions of the archival storage
ISO 16363 Link	Not explicitly mentioning the process of making copies, but section 5.1.2 comes close, where it mentions managing copies: “Repositories may require a	5.1.2

	different number of copies for each class, or manage versions needed to meet access requirements”	
--	---	--

Other frameworks and reference documents

Reference Document	Term used in framework to refer to the process	Section
OAIS Link	Replace Media Replication Refreshment	4.2.3.4 5.2.4.3 5.2.4.2
PREMIS Link	Replication, Media Migration, Media Refreshment	Glossary

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/93608373/Archival+Storage#ArchivalStorage-Redundancy ; https://wiki.tib.eu/confluence/spaces/lza/pages/93608943/Technical+infrastructure
	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-fe2023062157386 (section 3.2.4)
			https://urn.fi/urn:nbn:fi-fe2024051731943 (Annex 3, section 2.1.1)
Archivematica	Digital preservation system	English	https://archivematica.org/en/docs/archivematica-1.17/admin-manual/maintenance/maintenance/#data-backup