Object Management Reporting (CPP-013)

CPP-Identifier	CPP-013
CPP-Label	Object Management Reporting
Author	Johan Kylander
Contributors	Matthew Addis
Evaluators	Franziska Schwab, 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 delivers reporting to enable the effective management of *Objects*. This must include a wide range of functional, operational, and statistical reports and analytics.

Inputs and outputs

Input(s)		
Metadata		*Provenance metadata*
		Technical metadata
		Rights metadata
		Errors and Warnings
Output(s)		
Metadata		*Provenance metadata*
Documentation / guidance		Reports
		Statistical data

Definition and scope

Object management reporting is a process in which the TDA creates reports; outputs *Provenance metadata;* and produces statistical data on the content it preserves. The reporting is needed to enable effective management of *Objects*. Reporting is done for both the TDA itself, helping it in its preservation planning and as input for **Risk Mitigation** (CPP-012), and for the designated communities and stakeholders that produce and consume the data. Effective reporting enables a TDA to conduct digital preservation and to select the *Information packages* or *Objects* when performing preservation actions.

Reporting on the content includes reports and *Provenance metadata* from other CPPs that perform actions on digital content. For example, this includes processes that periodically check the quality of the data in **Data Quality Assessment** (CPP-019) or during triggered events such as **Ingest** (CPP-029) and **Enabling Access** (CPP-025):

- Reports from periodic **Integrity Checking** (CPP-003);
- Preservation actions taken and documented in **Data Corruption Management** (CPP-004);
- **Virus Scanning** reports (CPP-007);
- **File Format Identification** (CPP-008) reports;
- **File Format Validation** (CPP-010) reports;
- **File Migration** (CPP-014);
- Data on **Emulation and Rendering Tools** (CPP-015);
- *Metadata* from **Metadata Ingest and Management** (CPP-016);
- *Metadata* from **Rights Management** (CPP-020);
- *Metadata* from **Significant Properties Definition** (CPP-022);
- Preservation actions taken and documented in **File Repair** (CPP-027);

- Data from the **Creation of Derivatives** (**CPP-028**) process;
- *Metadata* from the access and use of *Objects* in TDA by the designated community;
- *Metadata* on the data volumes and growth rates of the *Objects* held by the TDA.

As the purpose of reporting is to enable the preservation and the management of digital *Objects*, the reports usually cover the main types of data that a TDA operates with. This may include data on:

- File formats;
- *Technical metadata* and significant properties of *Objects;*
- Tools and software needed to manage the content;
- The health of *Objects* and actions taken to remedy errors;
- The hardware infrastructure;
- *Metadata* and metadata standards;
- The provenance of the *Objects;*
- The licenses and rights to retain, preserve and provide access to *Objects;*
- The level of usage of *Objects* in the TDA;

The report types produced in the reporting process can be categorised as follows:

- *Statistical data* (e.g. number of *Objects*, file sizes etc.): Used for helping a TDA or stakeholder assess the scope of preservation actions;
- *Provenance metadata* (i.e. documented preservation actions by the TDA): Used for assessing the state of *Objects*, the so called "health" of the *Files*, and to add authenticity by documenting the life-cycle of all digital content in a TDA;
- *In depth machine-actionable reports* (usually produced by tools): These are used as input in other processes conducted by the TDA;
- *Links to knowledge base resources* (e.g. file format registries): These are used to help a TDA in preservation planning.

Reporting activities as well as the report types listed above feed into a wide range of management processes and decision making. Some examples include:

- Risk assessment and preservation planning (e.g. using reports on file formats);
- Reporting to funders and stakeholders (e.g. by providing usage statistics);
- Capacity planning and purchasing decisions (e.g. purchasing and refreshment of storage infrastructure based on statistics on data volumes and growth rates);
- Compliance assessments (e.g. using reports on rights and participant consent to check that personal data is retained in compliance with GDPR);
- Optimisation and cost management (e.g. identification of expensive or inefficient processes based on execution time and steps required, or optimising the use of different storage classes for frequently or infrequently accessed data based on usage statistics);
- Identification of anomalies in error rates or failures (e.g. unusually problematic content from specific depositors);
- Deaccessioning of *Objects* (e.g. reports on *Objects* that no longer need to be retained because their retention schedules are about to expire, or reports on quality assessment that mean *Objects* no longer meet the needs of a designated community);
- Environmental sustainability (e.g. use of storage, processing and access statistics as an input to calculating carbon footprint).

When the process description uses the term *report*, it can be any of the above.

Process description

Trigger event(s)

Trigger event	CPP-identifier
Ingest	CPP-029
Enabling Access	CPP-025
Risk Mitigation	CPP-012
Data Quality Assessment	CPP-019
Reporting needs from the stakeholders or designated communities	

Step-by-step description

No	Supplier	Input	Steps	Output	Customer
Analysis of Requirements					
1a	Many CPPs, like Ingest (CPP-029), Enabling Access (CPP-025), Risk Mitigation (CPP-012), Data Quality	Reporting request (from CPPs, stakeholders or the consumers) in the form of a Report Specification	The TDA receives the request and determines what data is required to provide the requested report	Specification of the data required for the report	

	Assessment (CPP-019)				
1b		Specification of the data required for the report	Determine if reporting data is available (e.g. is it already an output of other CPPs)	Availability of data required for the report	
1c		Availability of data required for the report.	If the reporting data is not available: (a) Determine the steps that would be needed in order to capture it (b) Discuss reporting requirements with the requestor and agree revisions to report	Modifications to (i) Report Specification and (ii) Specification of data required for the report	
2a		Specification of data required for the report (from 1a or 1c)	If needed, CPPs that produce the required data are revised so that the necessary data is produced and compiles the report (as provided by other CPPs) The TDA gathers and aggregates the data needed for the report.	Data required for report	
2b		Data required for report	The report is compiled and reviewed with the requestor.	Finalised Specification of Report	

		Report Specification	If further modifications are needed to the data and data collection, then go to step 2a.		
		Data Collection			
3a		Reporting schedule	If the report is required on a regular basis, e.g. monthly, then a reporting schedule is set and data is collected on a regular basis	Data required for report	Many CPPs, like Ingest (CPP-029), Enabling Access (CPP-025), Risk Mitigation (CPP-012), Data Quality Assessment (CPP-019)
3b		Data required for report (from 3a)	Gather and aggregate data for report	Data for report.	
		Production of a report			
3c		Data for report (from 3b)	Produce Report and deliver to Requestor.	Report	
		Finalised Report Specification (from 2b)	If a report is produced regularly, then loop to Step 3a.		

The steps above are grouped into three main stages:

Stage 1: Analyse report requirements and determine the availability/feasibility of collecting the required data.

Stage 2: Collect data (updating source CPPs if needed) and produce the first version of the report for review with the requestor.

Stage 3: Productionise report generation and produce the report on a regular basis (if needed).

Rationale(s)[1] and worst case(s)

Rationale	Impact of inaction or failure of the process
The TDA knows what *Information packages* it preserves and what they contain	Without being able to locate and report the data it holds, the TDA cannot perform any preservation actions or provide (e.g. searching by *Rights, Provenance* and *Descriptive metadata* within the *Information packages*).
The TDA knows which file formats it holds and can locate these *Objects* within *Information packages*	Without knowing in what format the *Objects* that a TDA holds are, the TDA cannot plan preservation actions or inform the designated communities about preservation strategies.
The TDA knows what metadata standards it holds and which standards apply to which *Objects*	Without being able to analyse the *Objects* and their significant properties, the TDA cannot plan preservation actions, provide discovery or access.
The TDA maintains and generates *Provenance metadata* during preservation	Processes that produce reports and provenance data must have their output stored. This information is needed to analyse the *Objects* in a TDA, create quality reports, provide designated communities with assurance of a high quality digital preservation
The TDA knows its hardware infrastructure and has a plan for its management	Without knowledge on the hardware infrastructure, the TDA cannot maintain high-quality bit-level preservation and report to the stakeholders.
The TDA knows the rights it has to preserve and provide access to its holdings.	Without knowledge of rights and retention schedules, the TDA cannot know with confidence that it has the permission to preserve its holdings or whether it is holding content that should be deaccessioned.

---

[1] Term derived from PREMIS.

2. Dependencies and relationships with other CPPs

Dependencies

CPP-ID	CPP-Title	Relationship description
CPP-003	Integrity Checking	Periodic integrity checking provides reports on the integrity of data and reports corrupted *AIPs.*
CPP-004	Data Corruption Management	Fixing corrupted *AIPs* produces *Provenance metadata* and data for quality reporting to the stakeholders.
CPP-005	Identifier Management	Soft dependency (i.e. may require): The management and reporting should require that the data is identified with PIDs.
CPP-007	Virus Scanning	Reports on virus scanning activities, frequency of threats, and outcomes of the actions provide essential input for operational management and risk assessment.
CPP-008	File Format Identification	File format identification reports are required for a TDA to enable it to manage its content.
CPP-010	File Format Validation	File format validation provide essential information on the well-formedness and validity of the *Objects*; validation errors; and data on the tools used in the process.
CPP-014	File Migration	File migration provides information on the outcome of the process as well as tools used.
CPP-015	Emulation and Rendering Tools	The process of selecting tools for emulation and rendering provides data to the TDA for reporting to the designated communities.
CPP-016	Metadata Ingest and Management	Metadata ingest provides new *Provenance metadata.*
CPP-022	Significant Properties Definition	To report on the characteristics of *Objects* for deeper analysis, the significant properties must have been defined.
CPP-027	File Repair	Fixing invalid *Files* produces *Provenance metadata* and data for quality reporting to the stakeholders.
CPP-028	Creation of Derivative Copies	Creation of new *Objects* provides data for statistical reporting.

Other relations

Relation	CPP-ID	CPP-Title	Relationship description
Affinity with	CPP-029	Ingest	Ingest is both an important provider of reporting data to the TDA (via other CPPs) as well as a customer, as the ingest checks and outcomes must be reported to the producer.
Affinity with	CPP-025	Enabling Access	Enabling Access of contents include providing *Provenance metadata*, statistical data and quality reports to the consumer.
Affinity with	CPP-012	Risk Mitigation	To plan and mitigate risks in preservation, a TDA needs to provide input on the preservation system, the quality of the data, significant properties in *Objects*, storage management *Metadata* etc.
Affinity with	CPP-019	Data Quality Assessment	To evaluate the quality of the data, the process needs input in the form of various *Metadata.*

3. Links to frameworks

Certification

Certification framework	Term used in framework to refer to the CPP	Section
CTS Link	/	Reporting is not defined directly as a requirement, but as the category "Digital Object Management", which contains the following requirements: Provenance and authenticity (R07) Deposit & Appraisal (R08) Preservation plan (R09) Quality Assurance (R10) Workflows (R11) Discovery and Identification (R12) Reuse (R13)
Nestor Seal Link	/	Reporting is not defined directly as a requirement, but there are specific requirements for various metadata categories, which are specified with

		reporting in mind: C27 Identification C28 Descriptive metadata C29 Structural metadata C30 Technical metadata C31 Logging the preservation measures C32 Administrative metadata
ISO 16363 Link	/	Reporting as such is not described, but referenced multiple times as a way for the TDA to demonstrate that it is meeting a requirement. See 3.2.1.3, 3.3.4, 3.5.1, 4.1.7, 4.6.2.1, 5.1.1.1, 5.1.1.3 and 5.1.1.3.1

Other frameworks and reference documents

Reference Document	Term used in framework to refer to the process	Section
OAIS Link	Generate Report	4.2.3.5
PREMIS Link	/	PREMIS does not dedicate any documentation to reporting, but it mentions reporting when highlighting the importance of storing events and metadata.

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://knowledge.exlibrisgroup.com/Rosetta/Training/Rosetta_Essentials/Data_Management/7.1_Searching_the_Rosetta_Permanent_Repository; and https://knowledge.exlibrisgroup.com/Rosetta/Knowledge_Articles/Is_there_a_list_of_all_of_the_SOLR_search_fields_in_Rosetta%3F ; and 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	English	https://digitalpreservation.fi/en/services/quality_reports/2024
Archivematica	Digital preservation system	English	Limited reporting. However, Archivematica is built on MySQL and ElasticSearch so there is the potential to generate more reports by directly accessing the databases/indexes.
Rosetta	Digital preservation system	English	https://knowledge.exlibrisgroup.com/Rosetta/Training/Rosetta_Essentials/Data_Management/7.1_Searching_the_Rosetta_Permanent_Repository; and https://knowledge.exlibrisgroup.com/Rosetta/Knowledge_Articles/Is_there_a_list_of_all_of_the_SOLR_search_fields_in_Rosetta%3F