Preservation of data integrity as part of the process routines

Content

Further information

Archival Storage

Specifications for submission information packages (SIP)

Technical Metadata

Structural Metadata

Preservation of data integrity as part of the process routines

The graphic below Integrity checks in the process organisation shows when in the course of an object's lifecycle checks are carried out to ensure data integrity and completeness, and how such checks have been implemented in the existing system processes. The headings numbers correspond to the event numbers in the graphic.

1. Completeness check in the acquisition process

When a member of the acquisition team receives a digital object, he or she conducts a completeness check as part of the quality control process. Before retrieving the object, the employee conducts a virus check.

2. Pre-ingest analysis

Checking for completeness is part of the pre-ingest analysis.

3. Uploading of objects in the source directory of the submission application

In consultation with the responsible acquisition team, the Digital Preservation team transfers the objects from the defined transfer directories and copies the objects to a server in the source directory of the submission application using a WinSCP client via an SSH-encrypted SFTP connection.

The SFTP standard contains internal mechanisms to check for integrity breaches during transfer.

The WinSCP client retains the original date and time stamps of the files.

4. Creation of Rosetta-compliant SIPs from defined transfer information packages

The transformation of transfer packages to Rosetta-compliant SIPs and AIP is described in the graphic Transformation of transfer information packages to SIPs and AIPs and the Automatic ingest process diagram.

The submission application, developed using the SDK provided by Ex Libris, converts the acquisition team's different data structures (transfer information packages) into Rosetta-compliant SIPs. The submission application creates Rosetta-compliant SIPs from various transfer information packages, and transfers them to Rosetta during the second step.

For connecting Leibniz Universität Hannover Institutional Repository to the digital archive via an OAI interface, TIB replicates the ZBW submission application. TIB undertakes the configuration of the submission application itself.

In the process of producing transfer packages, both submission applications create one MD5 checksum per file, which they store in an METS file. The names of all files belonging to the SIP are recorded in the structMap of the METS file.

The Rosetta-compliant SIPs created in this way are transferred to Rosetta when the deposit process is initiated.

5. Deposit: submission of objects to Rosetta

Once the deposit has been initiated, the Rosetta-compliant SIP runs through the validation stack. The following processes are carried out in the validation stack:

- Format identification using DROID
- Format validation using JHOVE
- Creation of three checksums per file; in the case of METS deposit, crosscheck of any checksums also delivered.
- Virus check
- Extraction of technical metadata using JHOVE, mediainfo or the NLNZ Metadata Extraction Tool
- Validation of the METS file

To check for completeness in the deposit process, the StructMap section of the METS file created by the submission application is checked.

6. Transformation of SIPs to AIPs

During the transformation process, the SIP is enriched with additional metadata and moved to various storage areas. Every time a transfer occurs, three new checksums are created and matched with those in storage; a completeness check is also carried out using the StructMap.

7.1 AIP update and integrity check as a process

Every time an AIP is updated, a copy of an IE is moved from permanent storage to operational storage. Every time a transfer occurs, checksums are recreated and matched with those in storage; a completeness check is also carried out using the StructMap. Integrity checks can also be initiated as a process within Rosetta, irrespective of any transfer.

7.2 Integrity assurance mechanisms of archival storage

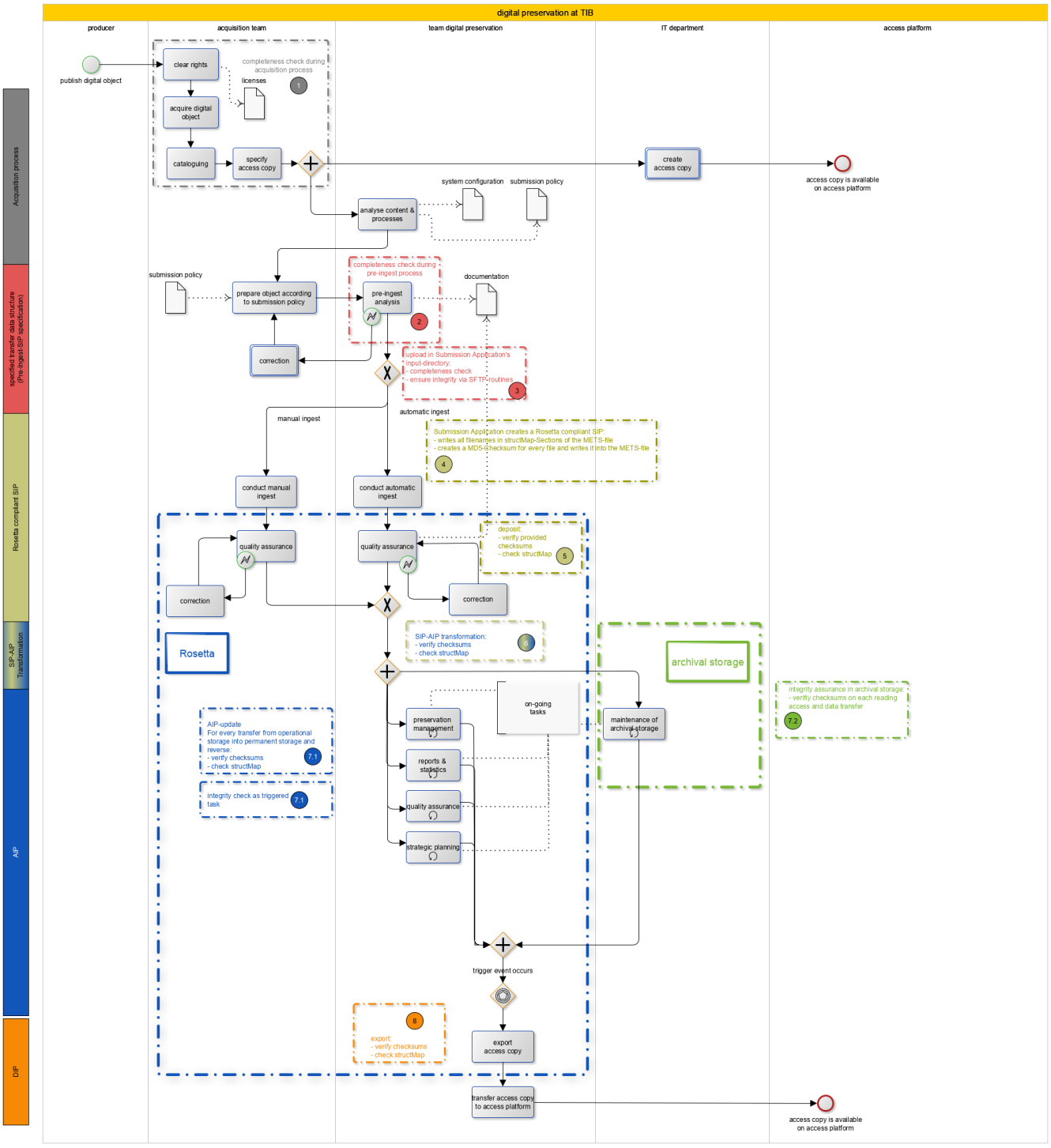
Archival storage's mechanisms for integrity assurance are described under Archival storage.

8. Export

Three checksums are stored in Rosetta for each file; during export, the checksums are recreated and matched. All files belonging to the representation are recorded in the ie.xml. This ensures that data integrity is safeguarded during export and that the files are complete.

If an error occurs during export, the process is terminated, and the system displays a relevant error message.

Integrity check in the process organisation


digital preservation at TIB
producer
acquisition team
team digital preservation
IT department
access platform
Acquisition process
specified transfer data structure (Pre-Ingest-SIP Specification)
Rosetta compliant SIP
SIP-AIP Transformation
AIP
DIP
publish digital object
clear rights
completeness check during acquisition process
licenses
1
acquire digital object
cataloguing
specify access copy
create access copy
access copy is available on access platform
system configuration
submission policy
analyse content & processes
submission policy
prepare object according to submission policy
completeness check during pre-ingest process
pre-ingest analysis
2
documentation
correction
upload in Submission Application's input-directory:
- completeness check
- ensure integrity via SFTP-routines
3
manual ingest
automatic ingest
Submission Application creates a Rosetta compliant SIP:
- writes all filenames in structMap-Sections of the METS-file
- creates a MD5-Checksum for every file and writes it into the METS-file
4
conduct manual ingest
conduct automatic ingest
quality assurance
quality assurance
deposit:
- verify provided checksums
- check structMap
5
correction
correction
Rosetta
SIP-AIP transformation:
- verify checksums
- check structMap
6
archival storage
on-going tasks
preservation management
maintenance of archival storage
integrity assurance in archival storage:
- verify checksums on each reading access and data transfer
7.2
AIP-update
For every transfer from operational storage into permanent storage and reverse:
- verify checksums
- check structMap
7.1
reports & statistics
integrity check as triggered task
7.1
quality assurance
strategic planning
trigger event occurs
8
export:
- verify checksums
- check structMap
export access copy
transfer access copy to access platform
access copy is available on access platform