
Submission Information Packages (SIPs)

Understanding SIPs

Deposit activities that Producer Agents submit to the Rosetta system consist of:

- Files
- Metadata about the files (such as creator, title, category, and subject)

Note

On some occasions, this data is part of a more complex object (such as datasets with various content items or whole journals with multiple issues). For such objects further information about their metadata their structure will be given as well

After a deposit activity is submitted, the Rosetta system processes the content as follows:

1. Files are organized into content intellectual entities (IEs). Depending on the material flow that the Producer Agent used to deposit content, either all of the files are stored in one IE, or a separate IE is created for each file. A content IE consists of:
 - **Files**, which contain the actual original data
 - **Representations**, which group files that represent different views of the same object. When content is deposited by a Producer Agent manually, a content IE can contain only one representation. When content is deposited automatically through FTP or NFS, representations can be organized pre-ingest in the METS file. For example, one representation may consist of files containing pages of a book as TIFFs, while another representation may consist of a single PDF as the entire book.
2. The Rosetta system aggregates descriptive metadata (such as title, author, and subject), which was provided by Producer Agents, and technical metadata (such as file size, file format, and MIME type), which was generated automatically, to the Metadata Encoding and Transmission Standard (METS). Each METS file represents a single IE.

Note

Descriptive metadata that do not have representations and files are considered structural IEs, which hold the metadata and structure of the complex object deposited (for example, the dataset metadata with the structure of its various content items or the journals and multiple issues metadata and structure). Descriptive metadata that have representations and file information are considered content IEs, which hold the actual digital content (for example, the various items under a dataset with their own metadata or the articles under the journal and issues). Structural IEs are optional when making a deposit. For more information, see the *Rosetta AIP Data Model* document.

3. All METS files (structural or content) representing IEs that were submitted within one deposit activity are grouped into a Submission Information Package (SIP) with the files. The METS XML file holds the aforementioned metadata along with the reference to the stream files that are deposited. (For more information on the structure of METS files, see [METS File Structure](#).)

METS File Structure

METS files contain information about intellectual entities (IEs), representations, and files. The table below describes the sections that a METS file contains.

METS File Sections

Section	Description
Descriptive metadata	Information provided by Producer Agents or staff users about the deposited content. This section can contain a reference to the metadata stored in an external content management system (CMS). Descriptive metadata is located at the level of IE, representation, and file.. The metadata is stored in the Dublin Core (DC) format.
Administrative metadata	Information that aggregates the following metadata: <ul style="list-style-type: none"> • Technical, which describes parameters of the deposited content, including file size, file format, and MIME type • Provenance, which describes parameters of users or processes that work with content, including the Producer Agent's name • Access rights, which define who can view content and when the content can be accessed • IE structural relationships that describe the structure of a structural IE Administrative metadata is located on the IE, representation, and file levels. The metadata is stored in the DPS Normalized XML format (DNX). For more information on DNX, see the <i>Rosetta Configuration Guide</i> .
Structural map	Hierarchy that defines how the IE's files can be logically grouped for easy navigation. A METS file can contain multiple structural maps that organize files by different criteria (for example, page scans can be grouped by page). Relevant only for content IEs.
File section	The <code><mets:fileSec></code> section that includes <code><mets:fileGrp></code> sections that contain the list of files grouped in a representation. Relevant only for content IEs. <ul style="list-style-type: none"> • Representation information: <ul style="list-style-type: none"> • USE – The usage of this Representation. In Rosetta it will be "View" even though METS allows more values such as "Thumbnail" or "TEI." • ID – The unique ID of the representation. • ADMID – The ID of the Administrative section that describes the representation. • File(s) information: <ul style="list-style-type: none"> • File ID – The unique ID of the file. • MIMETYPE – The file's Mime type (also described in the technical metadata section). • ADMID – The ID of the Administrative section that describes the file. • <code><mets:FLocat></code> - The file location element that points to the location of a content file. It uses the XLink reference syntax to provide linking information that indicates the actual location of the content file, along with other attributes specifying additional linking information. • <code><FLocat></code> is an empty element. The location of the resource pointed to MUST be stored in the <code>xlink:href</code> attribute.