Overview Guide

Version 7.3

CONFIDENTIAL INFORMATION

DISCLAIMER

TRADEMARKS

Document released: June 2022

Web address: http://www.exlibrisgroup.com

Table of Contents

1

Introducing the Rosetta System

The Rosetta system is designed to enable effective preservation of, and access to, digital heritage collections. With the Rosetta system, large amounts of digital data, including audio, video, and text content, can be stored and managed.

Rosetta enables various content providers to deposit content and make it available to library staff, the public, or to specific content consumers. To ensure the preservation of only high-quality, appropriate content, Rosetta provides various tools for managing content and content providers. Libraries have the option of storing material indefinitely or for a specified amount of time and then deleting it (tentatively or permanently).

Rosetta is a Web application that can be accessed using current Web browsers, such as Microsoft Internet Explorer, Mozilla Firefox, and Google Chrome. End-user interfaces such a library's Online Public Access Catalog (OPAC) or an application such as Primo or Voyager can be integrated with Rosetta to provide access to repository objects for external content consumers.

This guide describes the architecture of the Rosetta system, explains the types of system users, and provides a glossary of terms.

2

The Rosetta System Architecture

This section contains:

Understanding the Rosetta System Architecture

The following diagram illustrates the organization of the Rosetta system components and information flow.

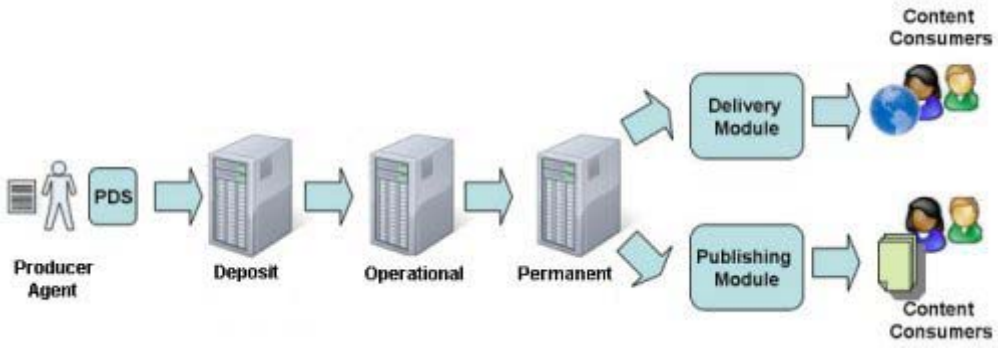

Figure 1: The Rosetta System Architecture Overview

NOTE:
Components can be managed on a single server or distributed among several servers. For more information, see **Storage Components** on page 8.

The information flow consists of the following stages:

1. Producer Agents log on to the Rosetta system through the Patron Directory Service (PDS).
2. Producer Agents upload files through the deposit server.
3. The Rosetta system moves the Producer Agents' content to the operational server, which processes content through the SIP Processing module.
4. During SIP processing, staff users review the content that Producer Agents deposited. They decide whether to approve the content, return it to Producer Agents for repairing issues ("reject" it), or decline it due to content issues.
5. Approved content is moved to the permanent repository. Returned and declined content is sent back to Producer Agents.
6. Both the Delivery Module and Publishing Module deliver the content from Rosetta through the interface to the content consumers.

Authentication Components: Patron Directory Service (PDS)

Users can be registered using either an external legacy application or the Rosetta system. The Patron Directory Service (PDS) enables the authentication and logon of users, regardless of the registration method. For example, the PDS can be configured to work with an external user database such as an LDAP directory service.

The PDS is configured by a System Administrator.

For detailed information and instructions on configuring a PDS, see the *Patron Directory Services Guide*, available under Cross-Product Information in the Documentation Center.

Storage Components

The following components store the content that Producer Agents deposit:

- **Deposit Server** on page **9**
- **Operational Server** on page **10**
- **Permanent Repository** on page **11**

Each storage component is characterized by the data it stores and the users who can work with this data.

IMPORTANT:

This guide refers to each component as a singular server. In fact, your library may store all components on one server or use a single server for each type of storage. References to a "server" such as an "operational server," therefore, may refer to a portion of a server or an entire server in a group of three or more servers.

For a description of server topology, see chapter 1 of the Rosetta *System Administration Guide*.

Deposit Server

The deposit server receives content from Producer Agents. Producer Agents can upload files of various formats and provide descriptive information about their content (such as title, author, and creation date).

The types and number of files, as well as the types of descriptive information that Producer Agents must provide, are defined by staff users. (For more information, see **Material Flows** on page **33**.)

Producer Agents use a Web-based wizard to deposit content.

Deposit Server Data

To logically organize content that Producer Agents deposit, the system groups Producer Agent files into deposit activities on the deposit server. (A deposit activity contains files that a Producer Agent has provided as one deposit.) In addition, deposit activities contain descriptive information (such as title, author, and creation date) that Producer Agents have provided about the content.

The deposit server stores the following types of deposit activities:

- Deposit activities that Producer Agents have saved as drafts for future deposits. Producer Agents can edit the draft deposit activities by adding new files, replacing the files, or editing descriptive information. When the draft deposit activity is ready, Producer Agents submit it for deposit.
- Deposit activities that staff users have returned to Producer Agents after review. Producer Agents can repair the issues and resubmit the deposit activity.
- Deposit activities that staff users have declined after review. Producer Agents cannot resubmit declined deposit activities.

Producer Agents manage their deposit activities through the Web-based interface.

From the deposit activities, the system creates SIPs and passes them on to the operational server.

Deposit Server Users

The following types of users work with the deposit server:

- Content providers:
  - **Producers** are responsible for providing content. (For more information, see **Producer Types** on page **26**.)
  - **Producer Agents** are responsible for depositing content. They self-register through the Web-based interface. (For more information, see **Content Providers** on page **24**.)
- Staff users:
  - **Deposit Managers** are responsible for configuring generic settings for Producers. The Producer Agents associated with these Producers use these settings when they deposit content. For example, Deposit Managers define the types and maximum size of files that associated Producer Agents can deposit.

    Deposit Managers are assigned and registered by a Back Office Administrator. (For more information, see the Deposit Managers section in the *Rosetta Staff User's guide*.)
  - **Negotiators** are responsible for tailoring the generic settings to the needs of specific Producers. For example, Negotiators can allow associated Producer Agents to deposit additional types of files.

    Negotiators are assigned and registered by a Back Office Administrator. (For more information, see Negotiators in the *Rosetta Staff User's guide*.)

Operational Server

After a deposit activity is submitted by a Producer Agent, the deposit server packages it as a SIP and notifies the operational server that this SIP is ready for processing. The SIP then goes through the automated phase of validation (validation stack), and then it becomes available to staff users. Staff users review the content and decide whether it needs to be approved, returned to the Producer Agent, or declined.

To access and review the content, staff users work with the Web-based interface.

Creating a SIP

To enable staff users to work with the Producer Agents' content, the Rosetta system moves the deposit activity to the operational server and converts each deposit activity into a Submission Information Package (SIP). (For more information, see **Submission Information Packages (SIPs)** on page **39**.)

Operational Server Users

The following types of staff users work with data stored on the operational server:

- **Assessors**, **Arrangers**, and **Approvers** are responsible for reviewing SIPs and deciding whether a SIP must be approved, rejected, or declined. Assessors, Arrangers, and Approvers are assigned and registered by a Back Office Administrator. (For more information, see Assessors, Arrangers, and Approvers in the *Rosetta Staff User's Guide*.)
- **Technical Analysts** are responsible for repairing technical issues that may occur with the SIPs. For example, Technical Analysts can manually assign a format to a file that couldn't be associated automatically. Technical Analysts are assigned and registered by a Back Office Administrator. (For more information, see Technical Analysts in the *Rosetta Staff User's Guide*.)

Permanent Repository

After a SIP is approved by staff users, the Rosetta system moves the intellectual entities (IEs) to the permanent repository. In the permanent repository, IEs are no longer grouped together as SIPs, though they retain, individually (as IEs), the metadata from the SIP.

From the permanent repository, the content can be delivered to content consumers through the Web and other channels.

Permanent Repository Data

The permanent repository is intended to store Producer Agent content that was approved by staff users for permanent preservation. As a result, IEs that are stored in the permanent repository cannot be updated, deleted, or rearranged.

When an IE must be changed (for example, its metadata or its format requires updating), the Rosetta system moves it back to the operational server. When the updating process is finished, the system returns the IE to the permanent repository and stores it as a new version of the IE.

Permanent Repository Users

The following types of users work with data stored in the permanent repository:

- **Editors** are responsible for managing content that was deposited by Producer Agents and approved by staff users for storage in the Rosetta system. (For more information, see Editors in the *Rosetta Staff User's Guide*.)
- **Content consumers** can search and view content in read-only mode. They access repository content through a public interface such as their library's online public access catalog (OPAC). The content that is available to them is defined by staff users (who define access rights options) and Producer Agents (who select specific options from those configured by staff users).

OAI Provider

Rosetta uses the Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH) for publishing IE descriptions to external systems. These systems (such as Primo) harvest Rosetta for records by calling the OAI Data Provider component.

This component is an integral part of Rosetta and cannot be configured or deactivated from the UI.

The OAI-PMH requests are expressed as HTTP requests. The base URL specifies the Internet host and port. The URL continues with a list of keyword arguments that take the form of `key=value` pairs. Arguments are separated by ampersands (&).

The OAI-PMH `key=value` pairs must use arguments that are supported by Rosetta.

The arguments supported are:

- `Verb = ListRecords`. This argument is mandatory.
- `metadataPrefix =` This argument is mandatory. The only supported values are `oai_dc` and `xepicur`.
- `set =` `set_spec` as it is defined in the Publishing Profile.
- `from` and `until` are optional.

Request example:

```
http:// <rosetta-server>/oaiprovider/
request?verb=ListRecords&set=books&metadataPrefix=oai_dc&from=20
10-01-01T08:05:04Z&until=2010-10-23T08:43:34Z
```

Delivery

The Rosetta system enables certain users to view content objects (such as intellectual entities, representations, and files) that are stored in the Rosetta system. These objects can be viewed by staff users (for example, Assessors can view the content that Producer Agents deposited) and external users (for example, a reader who has a subscription to the library). For the purpose of this guide, these users are called content consumers.

The following diagram illustrates the organization of the components that enable content delivery.

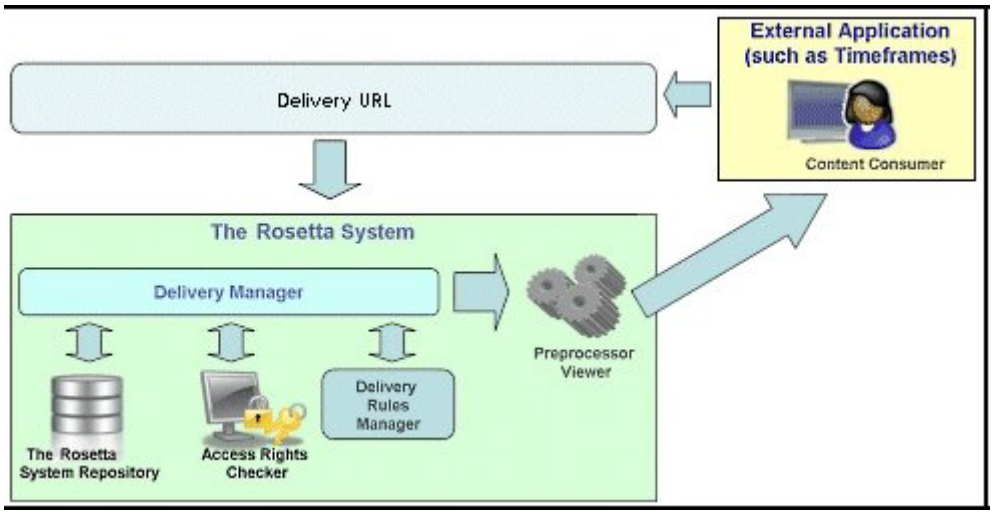

Figure 2: Delivery Information Flow

The delivery flow assumes the integration of an external viewer application such as a library's OPAC with Rosetta's delivery system. It consists of the following stages:

1. A content consumer uses an external application to request a content object from the Rosetta system.
2. The external application sends the request to the Delivery Manager module of the Rosetta system by using the Delivery URL. The requests are expressed as HTTP requests. The base URL specifies the Internet host and port. The URL continues with the PID of the requested IE or Representation (for example, `http:// <rosetta-server>/delivery/DeliveryManagerServlet?dps_pid=IE1077`. For more details about the Delivery URL and optional parameters, see the *Rosetta Configuration Guide*.
3. The Delivery Manager retrieves the content from the repository.
4. The Access Rights Checker determines whether the content consumer has the appropriate privileges to view the requested content object.

5 If the content consumer has the relevant access rights, the Delivery Rules Manager verifies the input parameters of the content object (for example, whether the object is an IE, representation, or file), and determines which viewing profiles, viewer preprocessor, and viewer must be used to display the content object.

6 The content object is forwarded to the Rosetta system viewer preprocessor, which then prepares the content for display in the viewer used by the external application.

7 When the viewer preprocessor finishes processing the content object, the Rosetta system sends the object to the external application viewer.

8 The viewer displays the content object to the content consumer.

For more information about delivery, see the chapter on Configuring Delivery in the *Rosetta Configuration Guide*.

3

Consortial Administration

This section contains:

Understanding Consortia, Institutions, and Departments

The Rosetta system enables multiple affiliated institutions to interact within one system and share resources.

To organize users, their content, and the way the users work with the system, the paradigm of a consortium is used. This paradigm organizes all Rosetta system users in the following entities:

- **Consortia**. Groups of organizations that manage the Rosetta system and allow their users to work with content. For example, a consortium can group various U.S. Presidential libraries.

NOTE:
The Rosetta system supports only one consortium per installation.

- **Institutions**. Organizations that a consortium contains. Each institution can have its own Producers, Producer Agents, and staff users. In addition, an institution can configure its own material flows and deposit control settings. For example, within the consortium of U.S. Presidential libraries, an institution can be represented by the Kennedy Presidential Library.

  Institutions are unique in that they exist outside of the consortial context.

- **Departments**. An institution's divisions or a dedicated group of people. For example, the Truman Presidential Library can include departments for government documents, historical documents, and private documents.

  In addition, each department has Editors who work with intellectual entities (IEs) within that department.

  The same department can exist within two different institutions because departments are referred to only within the context of an institution.

The diagram below shows the schematic representation of the consortium hierarchy:

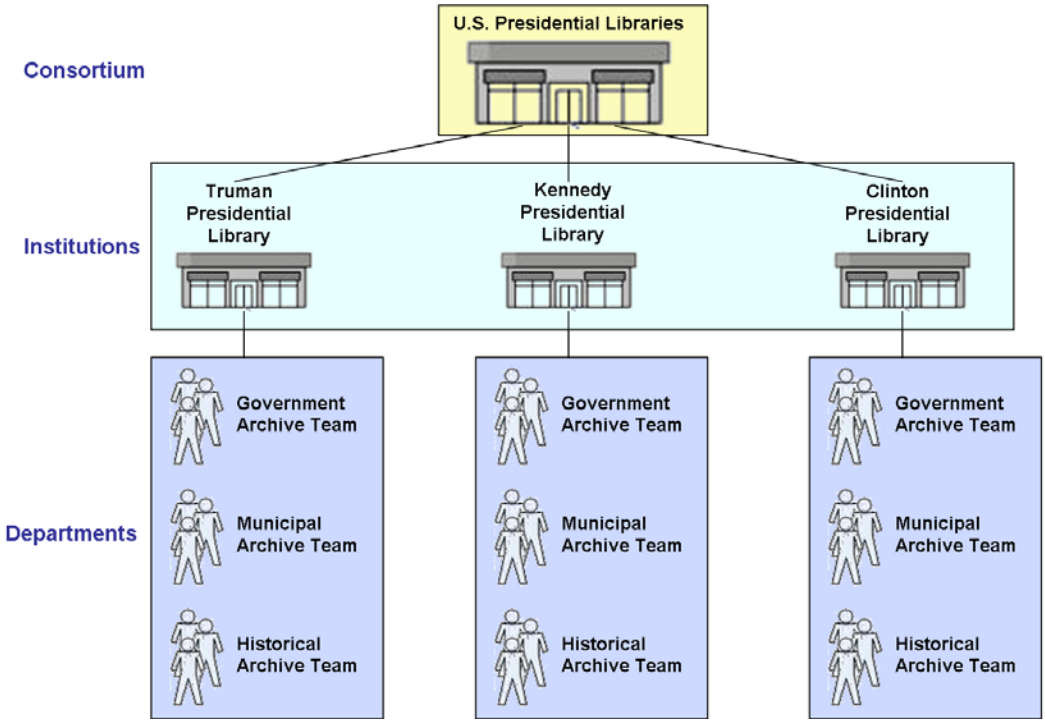

Figure 3: Sample Consortium Hierarchy

Configuration items such as Producer profiles and material flows are associated automatically with the institution of the user who created them.

Consortium Structure and the User Interface

The Rosetta user interface (UI) is organized by user roles that correspond to consortium-wide administration, on the one hand, and institution-specific

management on the other. An additional section of the UI is designated for individuals who deposit content. The UI breakdown is as follows:

Table 1. UI - Consortium - Role Association

UI Module	Consortial Range	User Role(s)
Administration	Consortium	Administrator
Management	Institution	Staff: Editor, Negotiator, Deposit Manager, Technical Analyst, Preservation Manager, Preservation Analyst.
Deposit	Institution	Producer, Producer Agent

Consortium-level Administration

Consortium details are predefined as part of the Rosetta installation. System Administrators can change the name and description of the consortium but not the code. They can also perform a number of consortium-wide functions shared by all of their institutions, including adding and editing member institutions. For information on adding institutions, see the Rosetta Configuration Guide.

Consortium-wide settings can be accessed by a System Administrator from the Administration module.

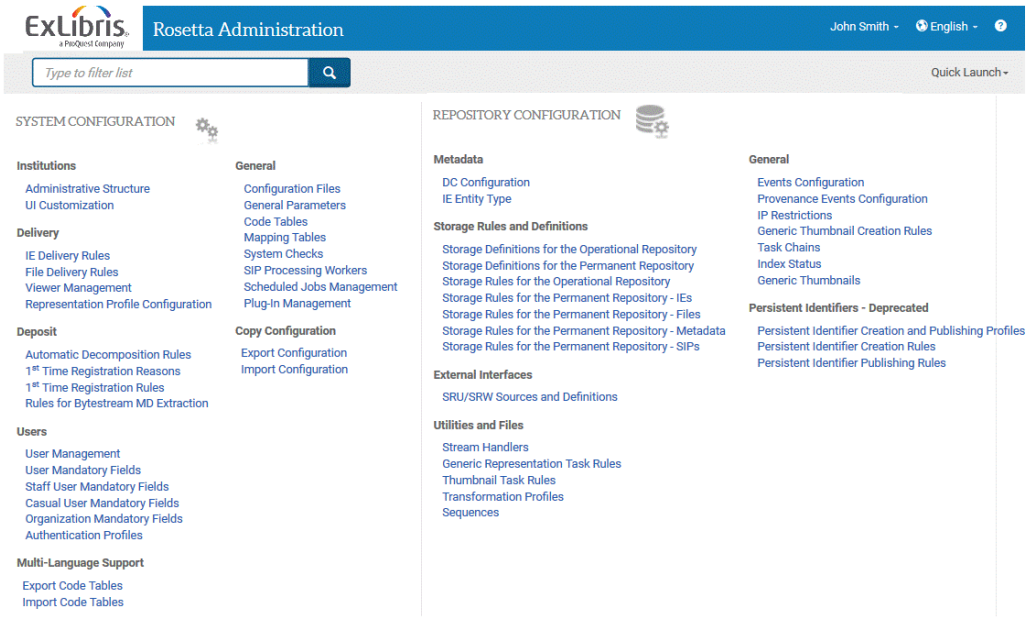

Figure 4: Administration/Consortium-wide Menu

Administrators handle consortium-wide items such as delivery and storage rules, task chains, and user management. Even if an institution is not making

use of the consortial features of Rosetta, the Administrator will be responsible for these areas of the system.

Institution-level Management

Staff users perform work on items that impact their institutions only. While consortium-level settings dictate rules that staff users must adhere to within the Management module, the management details themselves are defined by staff users like Deposit Managers, Negotiators, and Editors.

The management menu lists the areas where staff users perform their tasks.

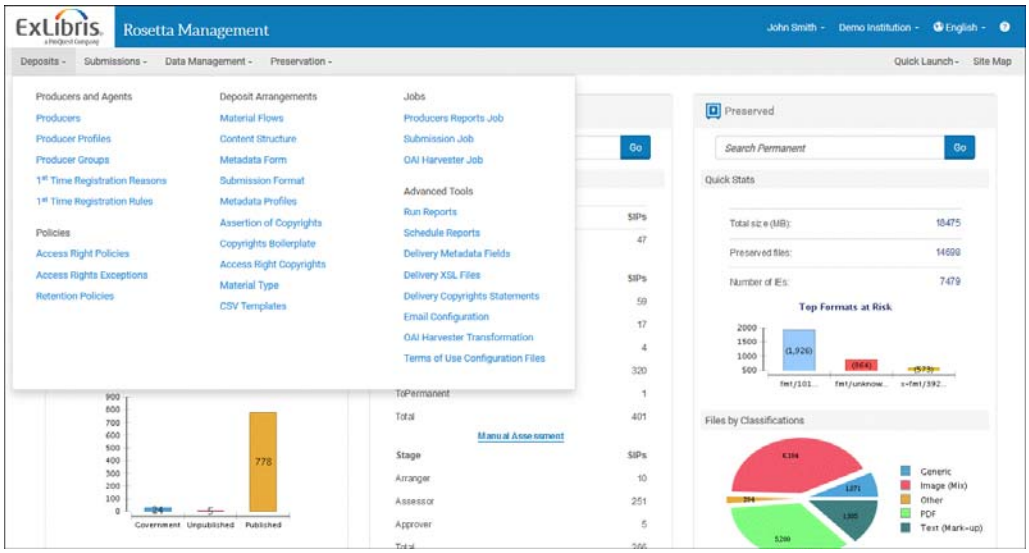

Figure 5: Staff Management Menu

Depending on their roles, staff users may enter one or more of the sections in the image above to perform their tasks.

A site map view of the Management module shows in detail all the areas staff users may access to perform their work.

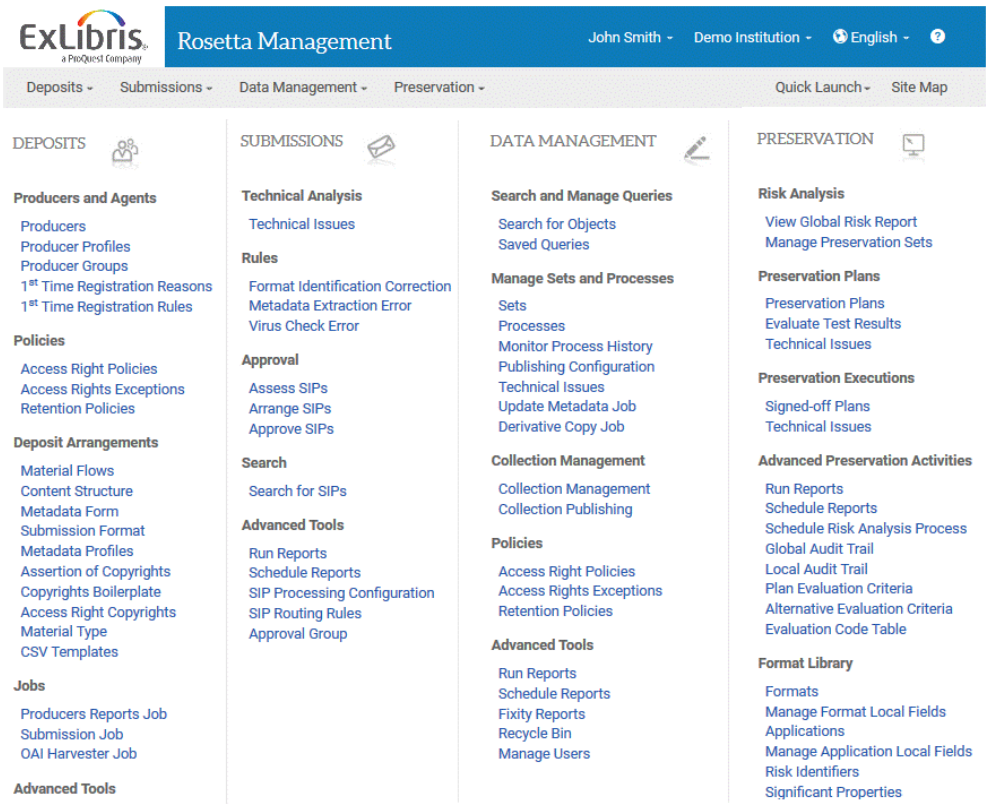

Figure 6: Site Map - Management Module

NOTE:

This site map shows links to all areas of the Management module, which are role-based as well as institution-specific. Users will see only those areas to which their role(s) have access. A Producer Manager, for example, will see the items in the DEPOSITS column but not the items in the SUBMISSIONS column.

Consortial Division of Roles and Tasks

Rosetta is organized according to types of user functions and tasks (or "items"). The table below lists these items and their distribution within the system.

Table 2. Consortial Scope

Item	Placement	Additional Information
Code tables	Mostly Administration/ consortium-level	Exceptions: Material Type, Producer Group, Approval Group, MD Forms Look-up, and Self-Registration Reason are institute-specific and belong to the Management module.
Delivery rules	Consortium-level	Decided by Administrator but staff can log on as different users to see different browser skins and outcomes.
Storage rules	Consortium-level	
Auto-decomposition rules	Consortium-level	Depends on the availability of dedicated tools.
Generic thumbnail generation rules	Consortium-level	Depends on the availability of dedicated tools.
Other rules (exceptions to the above)	Institution/ Management	Format identification auto-correction rules, files extension mismatch rules, metadata extraction error handling rules, SIP routing rules, SIP submission handling rules
Mapping tables	Consortium-level	
Configuration files	Consortium-level	
General parameters	Consortium-level	
Plug-ins	Consortium-level	After registration at the consortium-level, plug-ins are available to institutions
Task chains	Consortium-level	Institutions can specify parameters for a task that differ from those of another institution.
Validation stack	Consortium-level	All institutions use the same tools for these checks.
User management	Consortium and institution-level	Administrators and User Managers can manage user roles at the consortium and institution levels, respectively.
Producer management	Institution	Includes Producers, Producer Agents, profiles, groups
SIP processing	Consortium and institution-level	Routing rules and processing configurations are handled on an institutional level; task chains used by processing configurations are handled at the consortium level.

Table 2. Consortial Scope

Item	Placement	Additional Information
Deposit module	Consortium and institution-level	Access rights are shared metadata and handled consortium-wide; all other aspects (such as metadata forms, submission formats, and material flows) are handled within the institution(s).

Additional discussion of configuration roles and tasks are available in the *Rosetta Configuration Guide* (see the chapter on Advanced Configuration) and the *Rosetta Staff Guide* (see Configure Settings sections of the various sub-modules).

4

Rosetta Users

This section contains:

Understanding Rosetta Users

The following types of users interact with the Rosetta system:

- **Producers** provide the content (For more information, see **Content Providers** on page **24**.)
- **Producer Agents** deposit the content provided by Producers (For more information, see **Content Providers** on page **24**.)
- **Staff users** review the content deposited by Producer Agents, manage Producers and Producer Agents, and configure options available to them when they deposit content. There are several types of staff users:
  - Deposit Managers
  - Negotiators
  - Assessors
  - Arrangers
  - Approvers
  - Technical Analysts
  - Editors
  - Data Managers

- Preservation Analysts
- Preservation Managers

For more information, see **Staff Users** on page **27**.

- **Back Office Administrators** configure how the Rosetta system processes the content that Producer Agents deposit. (For more information, see **Back Office Administrator** on page **31**.)
- **Content consumers** access content deposited by Producer Agents. The content consumers have read-only access. (For more information, see **Content Consumers** on page **31**.)

Content Providers

Content is provided to the Rosetta system by Producers and Producer Agents. Both Producers and Producer Agents depend on the configuration and assignment of **Producer Profiles** and **Producer Types** to determine their roles and actions as content providers.

This section describes these concepts in the following subsections:

Producers are responsible for providing content for preservation in the Rosetta system. Producers can be represented by individuals (known as Individual Producers) or organizations (known as Group Producers).

Producers authorize individual users, known as Producer Agents. These Producer Agents are responsible for depositing and managing content. The Producer may assign multiple Producer Agents to submit content, as shown in the diagram below.

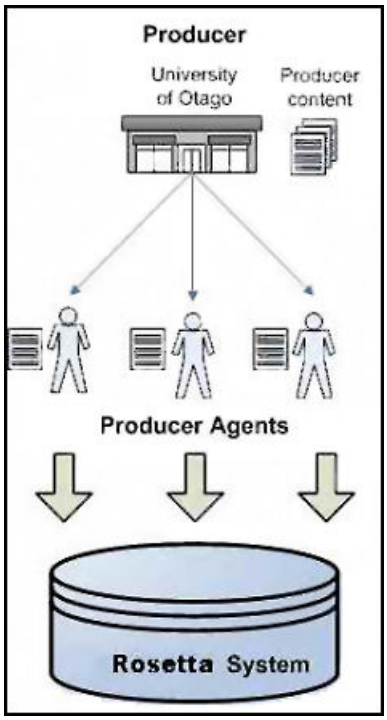

Figure 7: Producer with Multiple Agents

Similarly, a Producer Agent can deposit and manage content of multiple Producers, as shown in the following diagram.

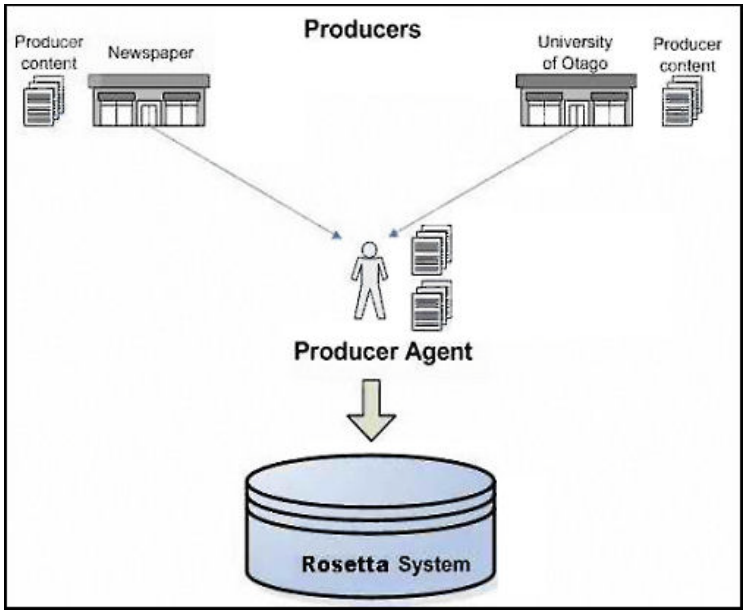

Figure 8: Producer Agent Depositing for Multiple Producers

Producer Agents deposit and manage Producer content through the Web-based interface.

Producer Profiles

Producer profiles govern how content can be deposited and how the Rosetta system processes this content. Producer profiles also define the amount of content that Producer Agents can deposit on behalf of Producers. Producer profiles are configured by staff users.

The Producer profile consists of the following components:

- **Deposit control settings**, which contain limitations on the amount of content that Producer Agents can deposit, and define whether the content needs to be reviewed by staff users.
- **Material flows**, which define the types of content that Producer Agents can deposit. This includes:
  - Information that Producer Agents must provide about the content
  - Access rights options available to Producer Agents
  - The deposit method (manual or automated) that the Producer Agent must follow

Producers can be associated with one of the following profile categories:

- **Generic**, which define material flows and deposit control settings that are assigned to Producers in a general, uniform way (unless personalized settings have been configured).

  Generic profiles are created by a Deposit Manager. (For more information, see Deposit Managers in the *Rosetta Staff User's Guide*.)
- **Personalized**, which can include customized generic material flows, additional material flows, and customized deposit control settings.

  Personalized profiles are created by a Negotiator. (For more information, see Negotiators in the *Rosetta Staff User's Guide*.)

Producers with a generic profile can contact a Negotiator and request that their profile be personalized, if required.

Producer Types

Producer types define how Producers are registered in the Rosetta system and how Producer Agents deposit content. The table below describes the Producer types that the system supports:

Table 3. Producer Types

Producer Type	Profile	Registration Method	Example
Registered	Generic	Self-registration through the Web-based interface	An individual or organization that regularly deposits content
Trusted	Personalized	Registration is performed by a Negotiator who can: ■ Upgrade an existing Registered Producer account ■ Create a new account	An organization that regularly deposits large amounts of content and requires personalized settings, such as additional types of content or increased available disk space
Internal (Staff users only)	Personalized	Registration by a Back Office Administrator	Any staff user

Staff Users

Staff users are responsible for managing Producers, Producer Agents, and the content that Producer Agents deposit.

The Rosetta system supports the following types of staff users:

- Deposit Managers (for more information, see **Deposit Managers** on page **28**)
- Negotiators (for more information, see **Negotiators** on page **28**)
- Assessors, Arrangers, and Approvers (for more information, see **Assessors, Arrangers, and Approvers** on page **29**)
- Technical Analysts (for more information, see **Technical Analysts** on page **29**)
- Editors (for more information, see **Editors** on page **30**)
- Data Managers (for more information, see **Data Managers** on page **30**)
- Preservation Analysts (for more information, see **Preservation Analysts** on page **31**)
- Preservation Managers (for more information, see **Preservation Managers** on page **31**)

Deposit Managers

Deposit Managers are responsible for configuring generic settings for Producers that provide content to the Rosetta system. These settings are automatically assigned to all Producers when they register in the Rosetta system.

Deposit Managers configure these settings by working with Producer profiles and their components (material flows and deposit control settings). Producer profiles define how the associated Producer Agents deposit content and how this content is processed.

Deposit Managers are responsible for:

- Managing Producer profiles, including adding, updating, activating, and deleting Producer profiles
- Managing deposit control settings
- Managing material flows:
  - Controlling material flows associated with a Producer profile, including associating, activating, and deleting material flows
  - Configuring material flow building blocks

Deposit Managers work with generic Producer profiles and generic material flows only. They cannot personalize Producer profiles or material flows for specific Producers. (For more information on personalizing Producer profiles and material flows, see **Negotiators** on page **28**.)

Deposit Managers are assigned by a Back Office Administrator who registers Deposit Managers and defines their privileges.

For more information, see Understanding Deposit Managers in the *Rosetta Staff User's Guide*.

Negotiators

Negotiators are responsible for working with Producers and tailoring the generic deposit configuration of the Rosetta system to the needs of specific Producers.

Negotiators tailor these settings by personalizing Producer profiles and their components (material flows and deposit control settings).

Negotiators are responsible for:

- Managing Producers, including adding, activating, and deleting Producers
- Managing Producer Agents, including adding and associating Producer Agents
- Personalizing Producer profiles
- Personalizing material flows

Negotiators are assigned by a Back Office Administrator who registers Negotiators and defines their privileges.

For more information, see Negotiators in the *Rosetta Staff User's Guide*.

Assessors, Arrangers, and Approvers

Assessors, Arrangers, and Approvers are responsible for reviewing submission information packages (SIPs) and deciding whether a SIP must be approved, returned to the Producer Agent, or declined. For more information, see **Submission Information Packages (SIPs)** on page **39**.

The following table shows the responsibilities of Assessors, Arrangers, and Approvers:

Table 4. Responsibilities of Assessors, Arrangers, and Approvers

Responsibility	Assessor	Arranger	Approver
Approving, returning, or declining the entire SIP	Yes	Yes	Yes
Approving, returning, or declining intellectual entities (IEs)	Yes	Yes	No
Rearranging IEs	Yes	Yes	No
Assigning a content management system (CMS) ID	Yes (Can assign a CMS ID to one IE)	Yes (Can assign a CMS ID to multiple IEs simultaneously)	No

Assessors, Arrangers, and Approvers are assigned by a Back Office Administrator who registers these staff users and defines their privileges.

For more information, see Assessors, Arrangers, and Approvers in the *Rosetta Staff User's Guide*.

Technical Analysts

Technical Analysts are responsible for handling technical problems (such as files of unidentifiable formats) that may occur with files that Producer Agents deposit.

Technical Analysts are responsible for resolving problems in the following stages of content processing:

- Deposit
- Loading
- Validation
- Bitstream extraction
- Enrichment
- Move to the permanent repository

Technical Analysts are assigned by a Back Office Administrator who registers Technical Analysts and defines their privileges.

For more information, see Technical Analysts in the *Rosetta Staff User's Guide*.

Editors

Editors are responsible for performing maintenance tasks such as:

- updating or editing metadata
- editing individual IEs
- adding representations to IEs

Editors are assigned by a Back Office Administrator who registers Editors and defines their privileges.

For more information, see Editors in the *Rosetta Staff User's Guide*.

Data Managers

Data Managers schedule processes, manage sets, and run activities that affect multiple IEs.

Data Managers are responsible for:

- Managing sets
- Scheduling and monitoring processes
- Publishing configurations, and
- Managing the recycle bin.

Data Managers can also work with individual IEs, but their primary role is to work on groups of data at an institutional level.

For more information, see Data Managers in the *Rosetta Staff User's Guide*.

Preservation Analysts

Preservation Analysts perform all tasks associated with risk analysis and loss prevention, including:

- populating Preservation libraries
- performing risk analyses on existing library collections
- creating, testing, and running plans and plan alternates

Preservation Managers

Preservation Managers perform all the tasks that Preservation Analysts do. In addition, they can sign off on preservation plans and reject preservation plans.

Administration Users

The Rosetta system is managed by the Back Office Administrator, who is responsible for configuring the Rosetta system. (For more information, see **Back Office Administrator**.)

Back Office Administrator

A Back Office Administrator is responsible for configuring the Rosetta system. This configuration includes defining how deposited content is processed, stored, and delivered. In addition, a Back Office Administrator defines how the Rosetta system users interact with the system.

For example, a Back Office Administrator configures SIP processing and storage rules.

For more information, see the *Rosetta Configuration Guide*.

Content Consumers

Content consumers can access Rosetta content only through an OPAC. They do not have direct access to items in the permanent repository. They can search the objects in the OPAC and view them by calling for a specific object using a deep link. The content that is available to content consumers is defined by staff users (who define access rights options) and Producer Agents (who select specific options from those configured by staff users).

5

Material Flows

This section contains:

Understanding Material Flows

Material flows define how Producer Agents deposit content. Staff users configure the following material flow parameters:

- **Metadata forms**, which define the types of information the Producer Agents must provide about the content. This can include author, title, owner, creation date, and description.
- **Submission formats**, which define how Producer Agents upload files and what limitations are applied to the files. Submission formats can be created for:
  - Manual deposit (For more information, see **Manual Deposit** on page **35**.)
  - Automated deposit (For more information, see **Automated Deposit** on page **37**.)
- **Access rights options**, which define the options the Producer Agents can select to control who can view the content.
- **Retention periods**, which allow users to limit the period for which content will be stored in the system.

- **Content structure templates**, which determine the XML framework for providing content metadata when content is deposited automatically. The Rosetta system supports Dublin Core and METS.

Once a material flow is configured, Deposit Managers can associate the material flow with generic Producer profiles.

Multiple material flows can be associated with a single Producer profile. For example, a Deposit Manager can associate the Producer profile with different material flows for depositing audio, video, and text content.

Similarly, a generic material flow can be associated with multiple Producer profiles. For example, a Deposit Manager can assign a material flow that enables depositing video content to multiple Producer profiles.

Negotiators can personalize generic material flows for the needs of specific Producers.

Metadata Forms

A metadata form contains fields that Producer Agents must complete in order to describe the content that they deposit.

Metadata forms are configured by Deposit Managers. To build metadata forms, Deposit Managers use the elements that are defined in the Dublin Core, an internationally recognized metadata standard. Dublin Core defines various metadata elements that describe content. These elements include:

- Creator
- Date
- Description
- Format
- Language
- Subject
- Title

When configuring a metadata form, Deposit Managers select the elements that need to be included on the metadata form. Optionally, Deposit Managers can specify possible values that Producer Agents can provide for each element. For example, a Deposit Manager can add an element for language and define three predefined options: English, German, and Spanish. A Producer Agent would then be required to select one of these languages when specifying the language of the deposited content.

Deposit Managers can associate a metadata form with multiple material flows.

Submission Formats

Submission formats govern how Producer Agents upload files and what limitations are applied to the files. Deposit Managers can associate a submission format with multiple material flows.

Submission formats define the following parameters of the Producer Agents' content:

- Types of files
- Number of files
- Size of files

Submission formats support the following methods of uploading files:

- **Manual Deposit** on page **35**
- **Automated Deposit** on page **37**

Manual Deposit

When depositing manually, Producer Agents select files to be uploaded. Producer Agents can upload files using one of the submission formats listed in the following table, as defined by a Deposit Manager.

Table 5. Submission Formats

Submission Format	Files Are Uploaded...	Limitations
HTTP Load	One by one	The submission format limits: ■ the total size and number of files that can be uploaded. ■ what file types can be uploaded. For example, the submission format can limit the deposit to: ■ JPG or PDF files only. ■ No more than 20 files. ■ No more than 5 MB for all files

Table 5. Submission Formats

Submission Format	Files Are Uploaded...	Limitations
Detailed	One by one	The submission format specifies ■ the exact number of each file type that must be uploaded. ■ the maximum size of each file to be uploaded. For example, the submission format can limit the deposit to one PDF file that is no more than 2 MB. In addition, the submission format can stipulate that the deposit must have two JPG files that total no more than 10 MB.
Bulk	Multiple files simultaneously	The submission format limits ■ the total size and number of files for the deposit, regardless of the file types. ■ what file types can be uploaded. For example, the submission format can limit the deposit to: ■ JPG or PDF files only. ■ no more than 20 files. ■ no more than 5 MB for all files.
CSV (comma-separated values)	One by one or multiple files	The submission format limits: ■ metadata delivery to a .csv file ■ metadata following a specific .csv template created by library staff ■ object delivery to one zip file

Automated Deposit

When depositing automatically, the files are uploaded to the Rosetta system from a predefined location on a remote server or local computer. The location is defined in the material flow. Producer Agents can use the automated deposit for uploading large amounts of content.

The following submission formats are supported for automated deposit:

- File Transfer Protocol (FTP), for content that is located on a remote server
- Network File System (NFS), for content that is located on a local computer

Access Rights Policies

An access rights policy is metadata that defines who can access an intellectual entity (IE) and when the IE can be viewed. Because staff users can assign an access rights policy to multiple IEs, this metadata is known as shared metadata. A Staff Member can also edit a shared access rights policy for a specific IE.

An access rights policy consists of expressions. An expression defines conditions that a user must meet in order to access the IE. These conditions might include an IP address range from which the IE can be viewed or a period of time after which the IE can be accessed. A policy can contain multiple expressions based on different criteria.

Access rights policies can be configured by a Back Office Administrator, a Deposit Manager, or a Negotiator. IEs are assigned rights based on the options that are given in the Material Flow to the Producer Agent.Then during SIP processing, Assessor/Approver/Arranger can re-assign different AR to each IE.

Once the IEs are in the Permanent repository, editors can re-assign AR policies to IEs.

Retention Periods

A retention period specifies a limit on how long content will be stored in a repository. Like an access rights policy, a retention policy can be assigned to multiple IEs and is thus known as shared metadata.

Retention periods can be based on time elapsed (for example, ten years following date saved to the system) or a specific date (for example, December 31, 2020). Material associated with the retention policy is deleted when the time elapsed or the date specified passes. Users can specify that the material be moved to recycling, and thus be available for restoring, rather than being deleted from the system permanently.

The system runs the retention period removal job nightly, along with a job that analyzes upcoming deletions, which can be viewed in a BIRT report.

Users can bypass the retention period if they want their content to be saved indefinitely.

Content Structures

Content structures determine the XML framework for providing content metadata. The Rosetta system supports various metadata standards, including Dublin Core, CSV, and METS.

Content structures are configured by a Back Office Administrator. Staff users can associate existing content structures with material flows.

6

Submission Information Packages (SIPs)

This section contains:

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** on page **40**.)

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.

Table 6. 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: ■ 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 *Rosetta Configuration Guide*.
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.

Table 6. METS File Sections

Section	Description

File section	The <mets:fileSec> section that includes <mets:fileGrp> sections that contain the list of files grouped in a representation. Relevant only for content IEs.
	
	This METS section contains information about all the files and some information about the representation.
	
	■ Representation information:
	  ■ 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:
	  ■ 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.
	  ■ <mets:FLocat> - 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.
	  ■ <FLocat> is an empty element. The location of the resource pointed to MUST be stored in the xlink:href attribute.

7

Common User Interface Elements

This section includes:

Web Interface

The Rosetta Web interface contains:

- core pages on which the main services are made available to users based on user roles and access rights
- administrative pages for setting up permissions and other variables during configuration
- common objects and links that appear across pages

This section describes some common links and objects that appear on many or all of the pages in the Rosetta interface. It also describes the Home pages for the Producer, Management, and Administration modules.

NOTE:
Rosetta does not support multiple instances in the same browser.

Home Page

Rosetta has three Home pages to accommodate each type of user and activity. After logging on to Rosetta, users see the Home page that is relevant to their role:

- **Deposit Home Page**
- **Management Home Page**
- **Administration Home Page**

Deposit Home Page

After logging on, Producers go directly to a list of My Producers so that they can select the Producer on whose behalf they are depositing. The Deposit module takes them through a wizard for submitting material.

Figure 9: Deposit Home Page

Management Home Page

The Rosetta Management Home page consists of a header and three columns containing key reports and search fields.

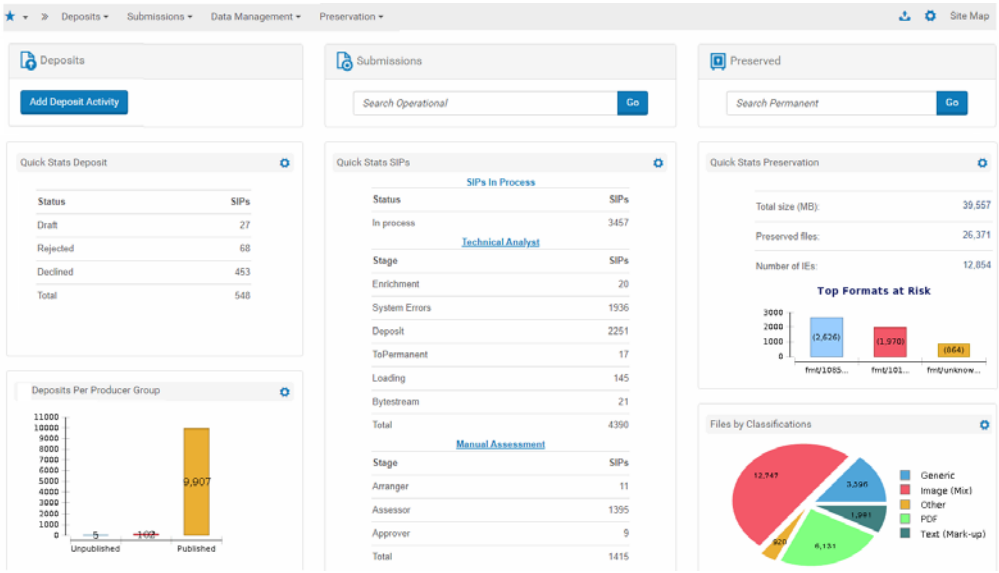

Figure 10: Home Page for the Management Module

Out-of-the-box, each column contains statistics, charts, and activities relevant to its area (Ingest/Deposit, Submissions, Preserved). The reports displayed can be configured by clicking the Configuration icon. For more information, see the *Rosetta Staff User's Guide* chapter, **Running and Viewing Reports** on page **42**.

Additionally, on this page you can perform searches by SID, PID, or text (see the *Staff User's Guide* for detailed information). You can also add a deposit activity, add personalized favorite destinations, or use the roll-over menus that cascade from the four segments (Deposits, Submissions, Data Management, and/or Preservation, depending on the user's role and privileges).

For details about searches and reports on this page, see the *Rosetta Staff User's Guide* chapter, **Searching and Reporting For All Staff** on page **23**

Administration Home Page

The Administration Home page contains a double-pane layout with a searchable flat list of sections and pages in the Administration module.

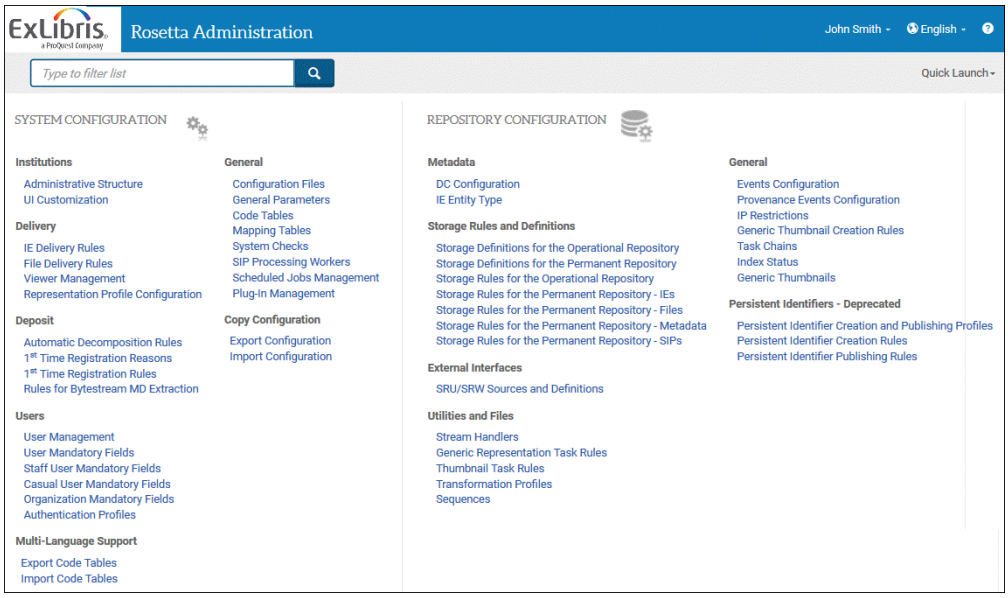

Figure 11: Administration Home Page

Header Features

Every page in Rosetta contains a header with the following features:

- Product logo, customizable (see Configuring Skins for Viewers in the *Rosetta Configuration Guide*)
- User name with a link to contact information and language (with option to change language)
- Navigation bar with links to main areas of the site. Areas change according to a user's role and responsibilities, such as whether the user is a Negotiator or an Administrator and whether they are using the Management module or the Administration module to carry out their tasks and responsibilities. Click the double arrow to open a search box that allows you to search for a menu option. Most navigation bars include a link to the Site Map.
- If a Management module also has Administrative privileges, then an icon  will be displayed, linking to the Administration module.
- A Favorites menu with links to your favorite management pages.

Features Repeated Across Pages

The following features are repeated across pages:

Site Map Link

A link to a role-based view of the site map displays on the right side of the header of most pages. Items shown on the map depend on the user's role. The figure below displays a site map containing all links for all Management-related roles.

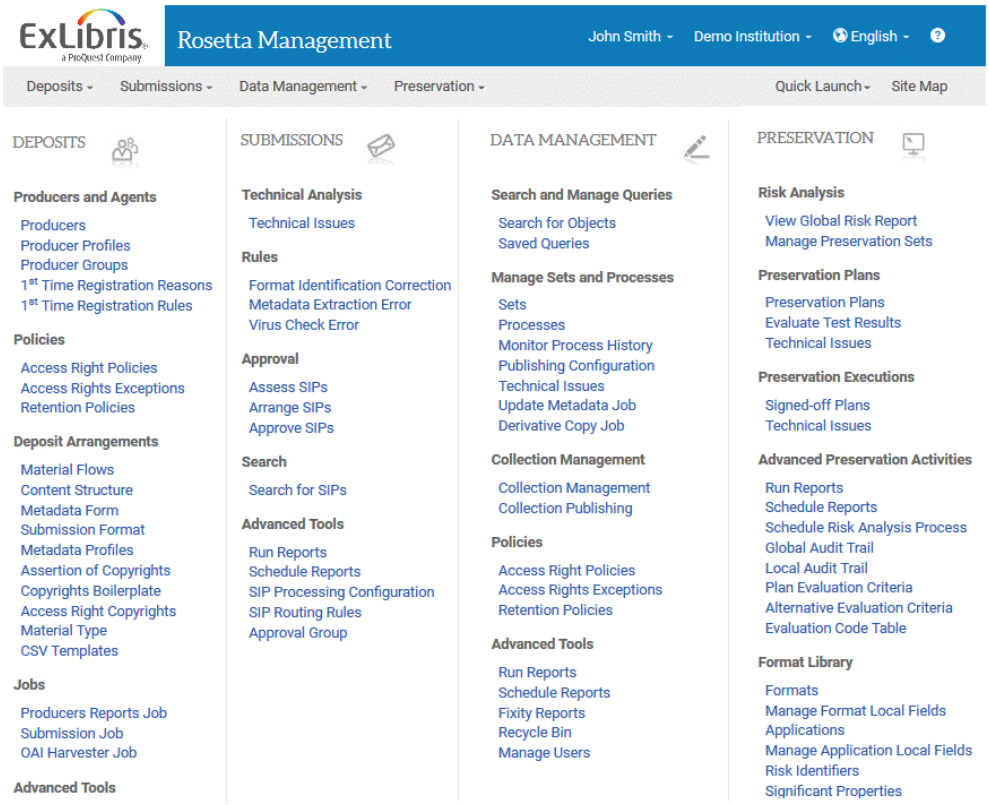

Figure 12: Site Map, All Management Roles

The links are the same as those that display from the navigation bar roll-over menus of Rosetta (see below).

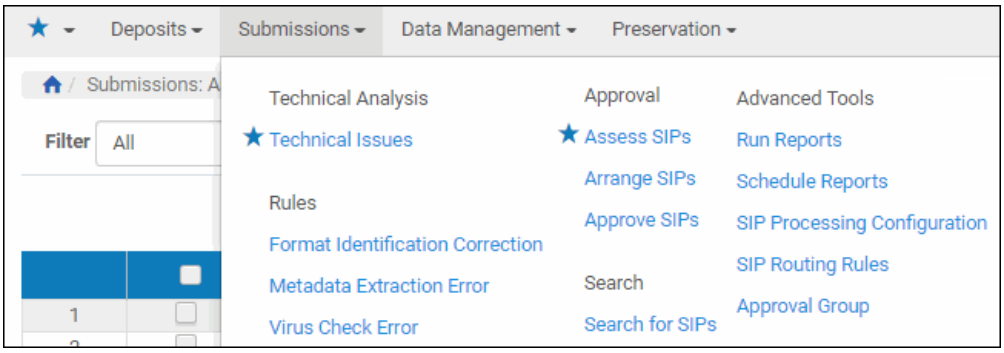

Figure 13: Submissions Menu

Select the star next to a menu option that you want to mark as a favorite. You can access your favorite menu options from the start icon at the top left of the Rosetta header.

Contact Link

On pages where an individual user's name appears (for example, the user detail page for a Producer Agent) or several individuals are listed (for example, a list of Producers), a contact icon displays beside each user's name. Clicking the icon opens a small pop-up window with contact information for the individual user.

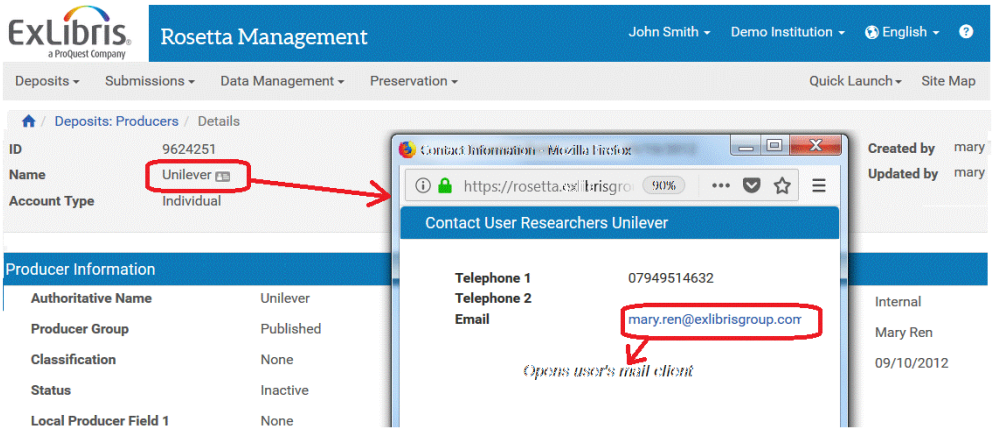

Figure 14: Contact Icon Opens User Information Form

Clicking the e-mail text link on the form opens a new blank message in the user's e-mail client. (If necessary, Rosetta launches the mail client application first.) The e-mail sender composes a correspondence as in a regular e-mail and sends it to the recipient.

NOTES:

- The e-mail client integration is supported for Microsoft Outlook.
- For reporting purposes, this feature gets logged as a *request* to send e-mail because the external application (Outlook) takes over the actual e-mail correspondence when it opens.
- This feature is assigned as a default privilege to a majority of Management roles: Assessor, Approver, Arranger, Technical Analyst, Negotiator, and User Manager.
- This feature does not overlap or interfere with the automatic e-mail template that opens when a Producer Agent's submission has been declined or rejected.

User Preferences

On several pages in Rosetta that display tables or lists of information, you can customize the way the information is displayed. Your settings are saved across all institutions.

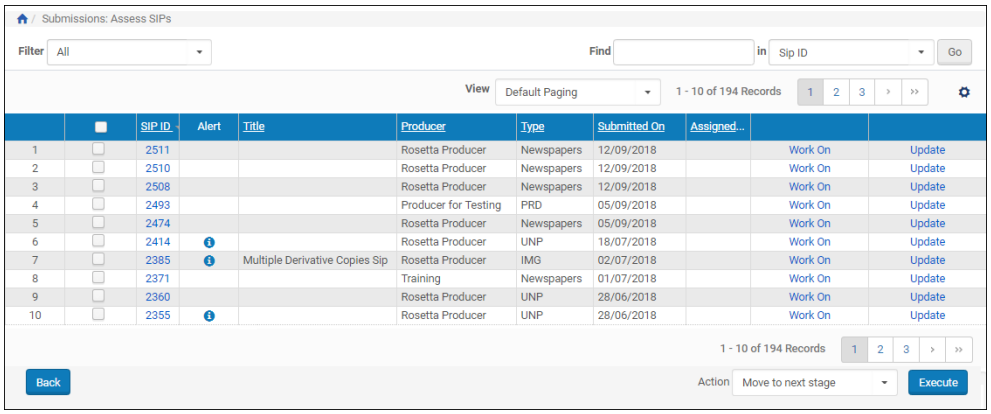

Figure 15: Assess SIPs

The following customizations are available:

- resize the columns of the table
- reorder the columns of the table by dragging and dropping
- from the Settings icon, you can select the columns that appear in the table
- display the number of records per page

8

Glossary

This section contains a glossary of terms used in the Rosetta system.

Access copy	When an Intellectual Entity (IE) contains files that cannot be viewed due to constraints in a library's technical environment, staff users can create a viewable copy called an Access Copy. Access Copies allow users to add representations of a derivative type to complete an IE.
Access rights	The rights of a user to view a particular object--and any related restrictions.
Access rights checker	A plug-in component that determines access restrictions for a user—from the user's particular context, and for the particular object.
Access rights exceptions	Exceptions to the object's access rights that may allow a certain user special access rights based on the user's or the object's characteristics.
Access rights options	Settings that define who can access the content deposited by Producer Agents, and when this content can be accessed. Access rights options are defined by staff users.
Access SDK	A code component that externalizes all the relevant APIs needed for creation of viewers and viewer preprocessors
Agent	A user or system that performs an action

Application Library	The Application Library contains all of the data regarding applications: name, ID, license end-date, and so forth. Each application can be related to one or more formats. The Application Library is managed globally (exposed to all installations of Rosetta), but some information is managed locally (for each installation). The connections between the Format Library and the Application Library are managed locally, but new applications as well as new formats should be added or removed in the global libraries.
Approval group	A group of Assessors, Arrangers, or Approvers to which the same group of SIPs is assigned by a back office Administrator. The group is based on business rules, and the parameters for calculating these values are the material type and classification. An approval group helps manage SIPs as they move through system processes.
Assessors, Arrangers, Approvers	Staff users responsible for reviewing SIPs and deciding whether a SIP must be approved, returned to the Producer Agent, or declined.
Audit event	An event that is stored as a discrete entry in an audit event table.
Back Office Administrator	An administration user responsible for configuring the overall framework that defines how Producers, Producer Agents, and staff users interact with the Rosetta system, and how content is processed.
BIRT	BIRT is an open source Eclipse-based reporting system that integrates with your Java/J2EE application to produce reports. Rosetta uses BIRT as its reports engine.
Bitstream	An object embedded within a bytestream that cannot be transformed into a standalone file without the addition of file structure (for example, headers).
Boiler Plate Statement	The copyright statement presented to a user when the user deposits items.

Bytestream	A compound file containing filestream(s) and/or embedded bitstreams.
Casual Producers	Casual Producers are associated with a single non-authenticated Producer Agent who submits material on a one-time basis through a special unrestricted URL. This URL links to a deposit interface limited to a set of submission wizards. Casual Producers are not managed actively. The option to allow creation of Casual Producers is configurable. If the parameter (in the General Parameters table) is set to false, there will be no option to create Casual Producers.
Classification group	Since significant properties are usually shared between multiple formats, the classification group is a way to aggregate these common properties so that they will be connected to all the relevant formats.
Consortia	The 2nd level of the consortium hierarchy. This is a group of institutions which allow them to share certain entities and support across institutional roles. The consortia level may not be used actively in any given installation but is always available and is integral to the structure of the system
Consortium hierarchy	This is a 4-level hierarchy is at which all roles or entities "live." The levels are: ■ Installation ■ Consortium ■ Institution ■ Department (a.k.a. Admin Unit)
Content structure	A specific type or structure of the submitted content that is supported by the Rosetta system. These are: ■ Set of files ■ METS and related files ■ Dublin Core and related files ■ CSV and related files.

Core repository	A two-layer Rosetta module. The upper layer, the services layer, is responsible for the management and execution of processes. The lower layer, the data layer, is responsible for storing and retrieving DE/IE. In the context of SIP processing, the core repository's upper layer is used (via API) for executing and monitoring task chains. The ability to execute and monitor task chains is important as some of the stage routines use task chains to implement the stage's processing instructions.
CSV	Comma-separated values. A file format used for storing database information in ASCII format (each entry or field is separated by a comma and each new row is represented by a new line).
CSV Loader	A Rosetta component that is used for loading SIPs data and metadata and creating a deposit activity from it.
Delivery	The component of the Rosetta system that enables content consumers to view content and content metadata that are stored in the permanent repository. Content consumers view the content and metadata using a Web browser and through different viewers that are suitable for the file's format. New viewers can be added to support different formats.
Delivery Manager	Accepts delivery request, checks access rights and directs user to appropriate viewer based on rules
Delivery Rules	The rules that determine how content is delivered to end users and external systems. Delivery rules are configurable by the customer and are analyzed in order by the Delivery Rules Manager, and when the system finds a matching rule, it does not continue the search.
Delivery Rules Manager	A component that determines the appropriate viewer for a particular delivery request.
Department (a.k.a. Admin Unit)	The fourth and lowest level of the consortium hierarchy which allows management of user roles and database collections directly.

Deposit activity	A transaction record stored in the Deposit Area for each deposit initiated through the Deposit Application. A Deposit Activity can have following statuses: ■ Draft ■ Submitted ■ Approved ■ Declined ■ Rejected SIP and deposit activity are synchronized through the SIP ID, which is returned by the staging server to the deposit application upon submission of the deposit.
Deposit application	The combined package of the deposit UI and the deposit server.
Deposit control settings	Settings that define the amount of content that Producer Agents can deposit and the amount of Producer Agent content that must be reviewed by the staff.
Deposit Managers	Staff users responsible for configuring generic settings for Producers (who provide content to the Rosetta system).
Deposit server	The server that receives the submitted deposit acquires the submission content, transforms it, and wraps it into standard METS format before forwarding the matching SIP XML record to the staging server.
Deposit UI	The Rosetta Web interface for submitting digital material to the repository. The client can replace the deposit Web interface component with a substitute interface, or bypass the component altogether and deposit directly into the deposit server using the Delivery APIs.
Deposit View	A deposit application UI instance with its particular look and branding, accessed through its own URL. A deposit interface is owned by one institution.
Descriptor file	A comma-separated values (CSV) file that holds information on the files, representations, and IEs used to create the new representation(s).

DNX	DNX metadata is implemented as an XML record, containing an unlimited list of sections (or "records") where each record contains unlimited numbers of attributes (or "properties"). Neither the record names nor the attribute names are limited by definition, so in theory they can hold any information. In practice, each DNX profile is limited by this flexibility and enables configurable validation, a read-only section, and default values. The motivation for defining another md-type called DNX comes from the need to collapse and aggregate administrative and technical metadata under one roof where the management, development, and viewing/ editing is much more simple.
DPS SDK	A set of programming tools (Software Development Kit) that allows programmers to develop specialized computer applications and adapt them to the Digital Preservation System (DPS).
DROID	A tool used by Rosetta for format Identification. For more information, see http://droid.sourceforge.net/.
Editor	The pre-defined role that grants a staff user privileges to interact with content in the repository. Three versions of the Editor role (View, Standard, and Full) determine the actions that a staff user may take with regard to sets and set members. Editors with sufficient privileges, for instance, can add new thumbnails for the images deposited by Producer Agents, edit the descriptive metadata provided by Producer Agents, and add new metadata.
Exchange package	The set of data used by the system for exporting and importing the files in the Add Representation environment. The exchange package contains the files themselves and a descriptor file with information about the IEs, representations, and files needed to create the new representation(s).
External user management	A model in which user information is managed through a third-party identity and access management system (IAM) or directory server. This data is used by the DPS.

File	A named and ordered sequence of bytes that an operating system can recognize. One or more files comprise a **Representation**.
Format Library	The Format Library contains all of the data regarding formats: name, description, related applications, related risks, and sustainability factors. Some of the information is managed globally (exposed to all installations of Rosetta) and some of it can be managed locally (stored within the local DB of the institution). In the local library, data can be added but not removed.
Gallery View	A SIP content view available to Assessors and Arrangers in which IE thumbnails display in addition to standard information.
Generic Producer profile	A Producer profile that defines material flows and deposit control settings that are assigned to all Producers (unless personalized settings were configured). Generic profiles are created by a Deposit Manager, and are assigned automatically by the Rosetta system when a Producer registers.
Group Producer	A Producer that is represented by an organization.
Human stage	A stage performed by a human agent (for example, Technical Analyst, Assessor, Approver) who uses the application's UI to manipulate the processed SIP. On entering a human stage, the application will forward the SIP's information to the appropriate agents and will mark the SIP as waiting for a response. The automatic processing continues once the appropriate human agent is finished working on the SIP.
Individual Producer	A Producer that is represented by an individual.
Installation	The first level of the consortium hierarchy, which encompasses all other levels.
Institutions	Independent work environments consisting of one or more administrative environments. More than one institution can comprise a system or consortium hierarchy.

Intellectual entity (IE)	A set of files that is considered a unit (for example, scanned pages of a book or a photograph). An IE is stored in a METS XML in the permanent repository. Alternatively, IEs can be structural IEs ,which represent complex hierarchal objects with relationships to other IEs. Structural IEs do not have any representations or files.
Internal Producer	Internal Producers are Trusted Producers (for use by staff) who can submit filestreams and/or descriptive metadata directly through the institution's Network File System (NFS).
Internal user management	Refers to a model whereby the user information is managed wholly within the DPS.
Itemized set	A group of objects whose set members are determined at the time the set is saved. There is no stored relation between the set members or the query from which they were derived.
JBPM	Java Business Processes Manager: a workflow and BPM engine that enables the creation and management of business processes that execute the stage routines and keep track of the SIP during its processing.
Material flow	Settings that define how Producer Agents can deposit content. Material flows are configured by Deposit Managers and Negotiators. A material flow can be associated with multiple Producer profiles. Similarly, multiple material flows can be associated with the same Producer profile. A material flow consists of: ■ **Metadata form** ■ **Submission format** ■ **Access rights options** ■ **Content structure**

Metadata	Information about the content that Producer Agents deposit. Metadata can contain both descriptive (such as author, title, and creation date) and technical (such as file size and location) information. Descriptive metadata is provided by Producer Agents. Technical metadata is automatically extracted from the content by the Rosetta system.
Metadata form	Contains fields that Producer Agents must complete in order to describe the content that they deposit. Metadata forms are configured by Deposit Managers and Negotiators. A metadata form can be associated with multiple Producer profiles. Similarly, multiple metadata forms can be associated with the same Producer profile.
Metadata management	A component of the work area that allows users to search, view, and edit metadata records as discrete objects; i.e., devoid of their association to any repository intellectual entity, representation, or file
METS	Stands for Metadata Encoding and Transmission Standard, METS is a common and widely use format. See: http://www.loc.gov/standards/mets/. The DPS staging server receives the data encoded in METS format. Deposit application is a DPS application used by Producer Agents to deposit digital materials. The deposit application is responsible for the file(s) acquisition and for the conversion from various content structures to a METS format. The Deposit Application is constructed from two main layers: the Deposit API layer and the Deposit Web interface.
Migration types	The only type of preservation alternative currently supported in Rosetta. The migration can be internal or external: ■ Internal - The converting tool is registered in Rosetta and is being activated as part of the process automation framework. ■ External – The conversion is performed outside of Rosetta and the preservation execution starts with receiving the converted files.

Negotiators	Staff users responsible for working with Producers and tailoring the generic deposit configuration of the Rosetta system to the needs of specific Producers.
OAI-PMH	The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH): A low-barrier mechanism for repository interoperability. Data providers are repositories that expose structured metadata via OAI-PMH. Service providers then make OAI-PMH service requests to harvest that metadata. OAI-PMH is a set of services that are invoked within HTTP.
OAIS	An Open Archival Information System (or OAIS) is an archive consisting of an organization of people and systems that has accepted the responsibility of preserving information and making it available for a designated community. OAIS is the ISO reference model for Open Archival Information System.
OTB	Out of the box: the configurations that come with installation and determine the default behavior of the system without alterations.
PDS	Patron Directory Service: A DPS Web component that facilitates user authentication and login in to the DPS. The PDS is part of the standard calling application package, but it is a distinct and separate component. It does not have a user database of its own. Rather, it can be configured to work with an institution's authentication server(s) and user database(s), such as an LDAP directory service.
permanent repository	The component of the Rosetta system that stores content that was deposited by Producer Agents and approved by staff users. From the permanent repository, the content can be delivered to content consumers through the Web and other channels.
Personalized Material Flow	A customized material flow prepared by a Negotiator for a specific Producer.

Personalized Producer profile	A Producer profile that includes customized generic material flows, additional material flows, and/or customized deposit control settings. Personalized profiles are created and assigned to Producers by a Negotiator.
Plug-in	Additional software that can be integrated into Rosetta and run by different modules. An example is a new type of technical metadata tool that can become part of the validation stack process.
PREMIS	Preservation Metadata: Implementation Strategies. An international working group charged with the following: ■ Define an implementable set of core preservation metadata elements, with broad applicability within the digital preservation community ■ Draft a data dictionary to support the core preservation metadata element set ■ Examine and evaluate alternative strategies for the encoding, storage, and management of preservation metadata within a digital preservation system, as well as for the exchange of preservation metadata among systems ■ Conduct pilot programs for testing the group's recommendations and best practices in a variety of systems settings ■ Explore opportunities for the cooperative creation and sharing of preservation metadata.
Preservation plan	The preservation plan is a structured workflow used by the Preservation Analyst (PA) to handle objects that are at risk. The workflow takes the PA through the stages of gathering documentation and general information, creating the preservation set, defining the suggested alternatives, running tests, and summarizing the results.
Preservation plan alternative	Each preservation plan should have one or more alternatives to ensure success. For example, the same plan may have two migration utilities that convert the source format to the target format. Each one of these is saved as an alternative, and the workflow allows the user to evaluate each utility and add information that is relevant for the utility being evaluated.

Preservation plan execution	After the institution signs off on a plan, the plan can be executed with no need to go through testing and defining the exact material. The plan's execution can be scheduled in advance or launched immediately
Preservation set	A preservation set is a set of intellectual entities (IEs) that is defined in the first stage of preservation planning. It starts as a logical set and becomes an itemized set every time a preservation plan is executed.
Pre-transformer	A routine that converts non-standard content structure into standard content structure, such that it can be transformed using the system's standard transformers. It is activated prior to the transformer as part of the SIP submission process.
Privileges	The discrete permissions that make up a user role. Privileges correspond to the interfaces and interface functions in the DPS.
Producers	An entity (person or organization) in whose name the digital material is submitted to the digital archive. A Producer is associated with one or many Producer Agents. Material flows and deposit parameters such as disk space quota and target group are assigned on the Producer level and apply to all associated Agents.
Producer Agent	A user who deposits digital material for the repository. A user may be associated with more than one Producer Agent role (typically staff depositing for the library as well as on behalf of a Producer).
Producer group	Every Producer is assigned to a Producer group by a staff user or by the system during the deposit registration process. These groups help the managing of Producers by allowing members to share aspects of their deposits, metadata, access rights, and other common characteristics.
Producer Manager	A user who has access to the Producer Management function (updating of Producer public fields and activation/deactivation of associated A.gents) in the Deposit Module

Producer profile	Governs how the associated Producer Agents can deposit content, and how this content is processed by the Rosetta system. Producer profiles also define the amount of content that Producer Agents can deposit. Producer profiles are configured by staff users.
Producer type	Defines how Producers are registered in the Rosetta system and how they deposit content.
PRONOM	An online registry of technical information about file formats. Created and managed by the UK National Archives.
Provenance	The documentation of the chain of events and actions (as well as related agents) that a specific object has undergone in the repository.
Provenance event	An action that involves at least one object or agent related to the repository.
Publishing	An extensible process that extracts and formats metadata for external uses.
Recycle Bin	A UI dedicated to the process of IE deletion from the repository. In this UI, a user with the necessary privileges can permanently delete or restore IEs that have been deleted by other staff users.
Registered Producers	Producers who are associated with authenticated Producer Agents, who have access to the full Deposit Module functionality (that is, they can review and track their deposits at any time). Registered Producers are assigned a generic set of material flows and therefore can start depositing material immediately upon self-registration, without any staff intervention.
Representation	The set of files, including structural metadata, needed for a complete rendition of an **Intellectual entity (IE)**. Each IE in a METS XML can contain multiple representations.

Retention Period	A period of time or an absolute date after which records stored in Rosetta can be deleted. Retention policies are defined for records that do not need to be stored indefinitely, such as papers required for legal purposes.
Risk analysis	A process that runs on the repository and outputs a list of files according to their Risk Identifiers.
Risk identifiers	Derived from either a query of the file attributes that put the files at risk (existing technical metadata) or a tool that extracts the technical metadata that describes the problem. Each file format can be related to one or more risk identifiers.
Role parameter	A role modifier or limiter that constrains the terms by which role privileges may be executed. For example, a Negotiator may be permitted to manipulate only those Producers in the Trusted Producer group.
Rosetta system	The Web-based software application designed to enable effective preservation of, and access to, digital heritage collections. With the Rosetta system, large amounts of digital data, including audio, video, and text content, can be stored and managed.
Sampling rate	Defines the amount of Producer Agent content that must be reviewed by staff users. The default sampling rate is defined by Deposit Managers, and the personalized sampling rate is defined by Negotiators.
Set	A physical list of objects. Sets can be created in two ways: ■ Itemized set: the user or system selects ("hand picks") the objects that will be included in the set. The result is a group of selected items. ■ Logical set: the user defines the set in search terms and lets the system generate the set at run-time. The result is a collection of objects that match the search criteria at the time the query is run.
Set management	A component of the Rosetta UI that allows authorized staff users to create and administer logical sets and itemized sets.

Set member	An IE, representation, or file that belongs to a logical set or itemized set.
SIP	A Submission Information Package (SIP) that is generated automatically by the Rosetta system when moving deposit activities from the Deposit Server to the Staging Server. A SIP consists of at least one of each of the following: ■ **Intellectual entity (IE)** ■ **Representation** ■ **File** SIPs contain information about provenance, location of the submission content, and the content structure. The DPS system defines an XML representation of a SIP (based on METS). The Staging Server receives an XML representation of a SIP from the deposit application and processes it.
SIP Items Tracking Table	A physical table in the Staging Server that contains information about the files that are associated with a single SIP. Each entry in the table points to a an entry in the SIP Tracking Table. The Items Tracking Table is used, among other things, for storing information about the processing of the SIP's files. Thus, for example, when a human agent processes the SIP, the agent's decision regarding each file (e.g., Reject, Decline, Accept, etc.) is stored in the appropriate Items Tracking Table entry.
SIP Processing	The process undergone by the SIP from the time it is received by the Staging Server until it is moved to the permanent repository. SIP processing resembles an assembly line: the SIP is automatically moved between processing stages according its processing configuration. During each stage, the SIP is processed according to the stage's predefined processing instructions. The final stage that indicates that the process has completed successfully is the move to the permanent repository.

SIP Processing Configuration	A set of stages, rules of flow between stages, and processing instructions managed in the JBPM. The processing configuration may vary between different types of SIPs, according to the predefined routing rules. When a SIP enters the Staging Server, the SIP routing officer decides, based on the predefined SIP Routing Rules, which SIP Processing Configuration applies to the SIP. Once the relevant SIP processing configuration has been identified, the SIP is processed according to the instructions found in the relevant configuration.
SIP Processing Stage	A logical step in the processing workflow. Each stage is composed of predefined logical processing instructions that are implemented by the stage routine and are executed by the application upon entering the stage. A workflow can be composed of a varying number of stages.
SIP processing state machine	This is the component that is responsible for determining what should be the SIP's next processing stage and for planting the appropriate processing instructions in the SIP tracking table. Knowing the SIP's current stage and the stage's result, the processing state machine can, by looking at the relevant processing configuration, determine what should be the SIP's next stage and which stage routine should be activated next. Once the SIP's next stage has been identified, the processing state machine asks the SIP tracking table manager to update the SIP's entry in the tracking table accordingly. The processing state machine can be thought of as a 'brain' that when told the SIP's current situation can read the appropriate processing configuration manual, determine what should be done next with the SIP, and store the processing instructions for the SIP's next stage in the appropriate place
SIP processing workflow	The process that handles the SIP from the point of submission to the staging server to the point that it moves to the permanent repository. The SIP processing workflow includes the following phases: ■ Transform the METS file to AIP ■ Validation stack ■ Approval ■ Enrichment ■ Move to permanent repository

SIP processing workflow configuration	A specific set of stages, rules of flow between the stages, and processing instructions. Processing configuration may vary between different types of SIPs, according to the predefined routing rules. When a SIP enters the Staging Server, the SIP routing officer decides, based on the predefined SIP routing rules, which SIP processing configuration applies to the SIP. Once the relevant SIP processing configuration has been identified (and linked to the SIP's entry in the SIP tracking table), the SIP is processed according to the instructions found in the relevant configuration. The SIP processing statemachine uses the relevant SIP Processing Configuration as a manual describing how the specific SIP should be handled throughout its processing.
SIP routing officer	A component that identifies, based on the predefined SIP routing rules, the SIP processing configuration that applies to a specific SIP.
SIP Routing Rules	A set of parameter-based rules that point out the appropriate SIP processing configuration for a combination of SIP attribute values.
SIP worker	An application program (thread) that executes the processing instructions associated with the assigned SIP's current stage.
Skin	A definition of the look of Rosetta pages. In this document, skin refers to colors, logos, and icons presented in the different Delivery viewers.
Staff users	Users who are responsible for managing Producers, Producer Agents, and the content that Producer Agents deposit. The following staff users exist: ■ **Deposit Managers** ■ **Negotiators** ■ **Assessors, Arrangers, Approvers** ■ **Technical Analysts** ■ **Editor**

Stage routine	Implements the processing instructions (internal or task chain) of a stage within a processing configuration. When a SIP enters a processing stage and is selected for processing by a SIP worker, the SIP worker executes the appropriate stage routine according to the instructions it receives from the SIP tracking table manager.
Operation module	The component of the Rosetta system that stores content submitted by Producer Agents. In the Operational module, staff users review the content and decide whether to approve it for permanent storage, return it to the Producer Agent, or decline it.
Statistics event	An event which is a calculated aggregate of events over a period of time, used for determining event measures (such as average or number).
Structural map	Deposited METS can contain structural maps for the entire IE, a group of representations, or a single representation. For the entire IE: comes from the deposited METS that does not reference the ID of any of the METS' file groups. It is linked to the IE, preserved, and can be exported. However, delivery and other operations may not be available for this type of structural map. For a single representation: a structural map within the deposited METS that shares an ID with a single fileGrp. For multiple representations (i.e., shared structural map): exists within the deposited METS that references the IDs of several file groups within the METS. The syntax for multiple references is: ID=”fileGrp_1 Id; fileGrp_2 Id; …” In the case of a shared structural map, the Staging Server is responsible for creating a copy of the structural map for each of the relevant representations upon loading the SIP to the Staging Server (since in the repository, every representation is self-contained and contains its structural map).
Submission content	The files and metadata prepared by a Producer for submission. Submission content should follow valid DPS content structure.

Submission format	Settings that govern how Producer Agents upload files and what limitations are applied to these files. Submission formats are configured by Deposit Managers (when a generic submission format must be created) and Negotiators (when a personalized submission format must be created). A submission format can be associated with multiple Producer profiles. Similarly, multiple submission formats can be associated with the same Producer profile.
Submission Information Package	See **SIP**.
System Administrator	An administration user responsible for configuring both the server on which the Rosetta system is installed and the Oracle database that Rosetta uses to store Rosetta-related data.
System stage	A stage performed automatically by the system, without receiving human input. The Stage routine implementing a system stage can either contain the code instructions of the stage or point to the task chain that contains the relevant code instructions.
Task	A program that performs an operation on an object within the Core Repository.
Task chain	An ordered list of tasks.
Technical Analysts	Staff users responsible for handling technical problems (such as corrupted files or files infected by a virus) that may occur with files that Producer Agents deposit.
Transformer	A program that converts standard content structure into a SIP METS format.
Tree view	A SIP Contents view available to Arrangers, in which the original tree structure of the files in the SIP displays in addition to the regular information.

Trusted Producers	Trusted Producers are Registered Producers who have negotiated access to personalized material flows. The level of personalization can go from the definition of default values for selected form fields to the automation of the deposit process by allowing submission of both file streams and associated descriptive metadata as files on an FTP server.
User	An individual or organization that interacts in some way with the system. A user may be a staff member or a patron that logs in to a module and uses the system or a user may be an organization in a more general sense. A user of the system can be assigned various roles such as Negotiator, Approver, and/or Technical Analyst. Some roles are limited to staff users, whereas others such as Producer Agent can be assigned both to staff users and patrons.
User role	A named group of privileges that a user is authorized to perform. Roles are based on expected workflows and job responsibilities in the DPS. They are fixed and not editable by the library.
User role profile	Comprises both a role and the role's associated parameters. Although role privileges are fixed, parameters will vary depending on the user. One user linked to the Negotiator - Full role may be assigned the parameter of Trusted Producer Group only, while a second Negotiator - Full may be assigned the parameter of All Producer Groups.
Viewer	An extensible component that handles the viewing of content.
Viewer pre-processor	An extensible component that is activated based on the delivery rules to prepare an object for viewing and redirect user to the relevant viewer.

Index

T

U

W

X