



### **Repository System. (V.2004-DC)**

This document describes the architecture of the repository "OCTOPUS".

OCTOPUS is an application designed as a centralized system for the management and archiving multiformed digital documents (collection) produced and managed autonomously by several institutions (authority).

#### **System architecture.**

The architecture of the system is based on three components:

1. The data archive ("Repository")
  - Provides document storage
    - i. Bibliographic description (text format)
    - ii. Multiform document files (MIME)
2. The "admin-user" application interface
  - a. Make available all the services provided by the system to both the end user and the administrator via Web-Based Interface.
    - i. Submitting Documents
    - ii. Administration and control (for each authority)
    - iii. Free Search (on End User Options)
    - iv. Search based on predefined lists
    - v. Edit / Correct Documents
    - vi. Managing access rights for each single document
    - vii. Free internet access to anyone up to "*bibliographic description*"

- viii. Control the access rights (view and download file) of the "body" for each single document
- 3. The relational database (MySQL)
  - a. Provides the indexing of various elements of the "bibliographic description" of the document. Speed up searching and accessing a document or set of documents, as well as simplifying the production of lists and document clusters for:
    - i. Year of Creation
    - ii. Typology
    - iii. Author
    - iv. Collection & Authority:
  - b. Keeps information about collections and other related to access control by the designated administrator for each institution:
    - i. Administrator access rights
    - ii. Information on collection & authority

## **The "Repository"**

Logical model of the document.

An essential element in the OCTOPUS architecture is the notion of a document, a logical abstraction that incorporates the following concepts.

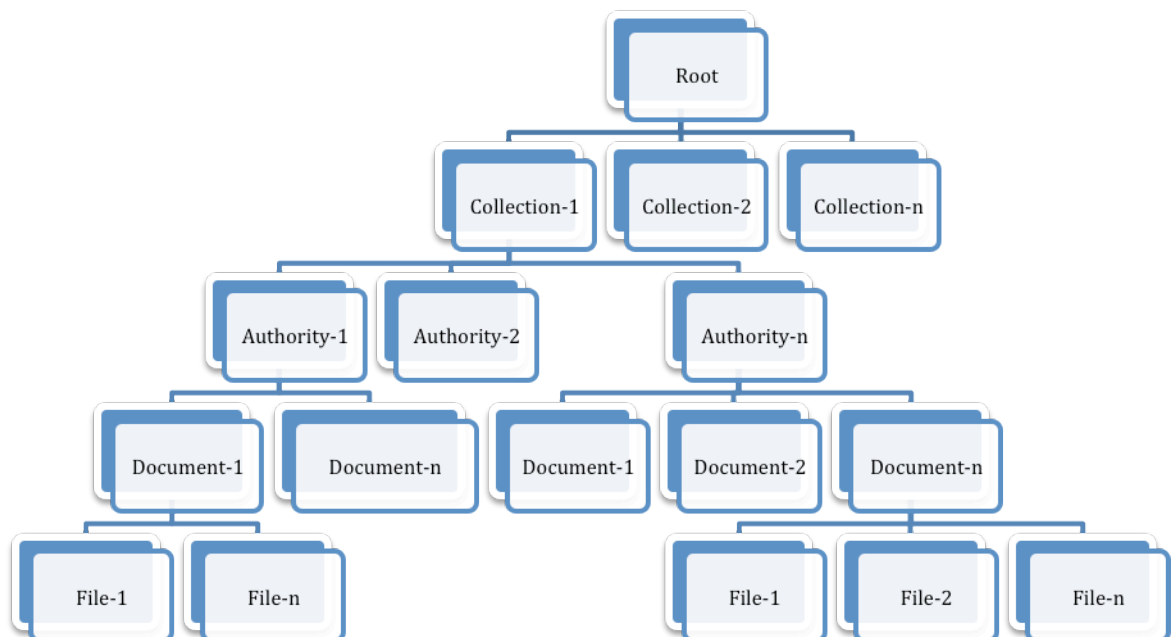
- Each document has a unique global name that is defined using the specific procedure described in "document naming".
- A document consists of a number of components. The two essential components are the "bibliographic description" and the "body" of the document.
- The bibliographic component is available in text format (ISO-8859-1), the "body" component is available in one or more formats (MIME). Exceptionally the "body" component may not be present in specific collection and
- The "body" component of a given format can also be structured on multiple files and sub-directories. For example, the "body" available in HTML can be structured into pages and organized so as to contain sub-directories.

## Physical organization of the repository

Physically, the "repository" is a portion of disk with defined access rights, not directly accessible from the "http" protocol and structured according to the "File System" model where, starting from a certain folder (Windows world) or "Directory" (UNIX world),

For each collection there is a subfolder / directory for each authority defined within the collection. Each single authority hosts a subfolder / directory for each document that belongs to it.

The sub-folder / directory of each document hosts at least two elements (files), the first defined "*bibliographic description*"; The second defined "*body*" can in turn be a simple "file" (MIME) or a set of files also structured in a further subfolder / directory.



### *Document naming*

Each document is assigned a unique code in the built-in system:

- Collection code
- Authority code
- Year of reference (intellectual creation of the document)
- Type document signature
- Sequential number

Typical code: Code\_collection1/Code\_authority1/year-Code\_type\_doc-number

## **The "administrator - user" application interface**

The application interface is provided by the OCTOPUS system. The application implemented in PHP language enables the dynamic construction of WEB pages to activate via the "menu" the corresponding services that allow you to interact with the system and produce dynamic answer pages to queries or actions required by the user user of the system.

The interface is mainly focused on three types of users:

- The administrator user (the librarian of the registered authority)
- The provider / owner of the data (single person or staff of registered authority)
- The common / final user (or simple internet browser)

## **Services**

The basic services of the system are organized into modules that interweave each other for the performance of the features that the system offers.

The basic services of the system are organized into modules that interweave each other for the performance of the features that the system offers.

### **1. Administration:**

A service reserved for registered users as administrators that allows you to carry out the administrative functions provided for all authority. Each single authority may register its own administrator; access to the service is permitted through authentication via UserName & Password.

- a. The Service allows user to:
  - i. Administer documents entered by (authorized or not) users
  - ii. Validation of the bibliographic document
  - iii. Identification code assignment
  - iv. Insert the document into the "Repository" and the database
- b. Editing an existing document in case of errors detected after it was inserted
- c. Remove a document if no longer relevant

### **2. Entry data:**

Service reserved for users of registered institutions or open to all.

The service is available after user authentication that can be done in two different ways. The service allows:

- i. To assist the user in compiling a set of forms designed to create the "bibliographic description" that characterizes the document that the user intends to submit to the system.
- ii. At the completion of the compilation and verifying of the entered data, the system sends a confirmation email of the sending to the e-mail address that the user has inserted in the

form and the administrator of the specification Institution because it provides for the administration phase.

3. Free Search:
  - a. Service available to any Internet user. The result of the research is the display of the "bibliographic description" that allows you to view / download the "body" of the document based on the rights / access modes attributed to the document itself.
4. Search based on pre-defined lists:
  - a. Service available to any Internet user. The result of the research is the display of the "bibliographic description" that allows you to view / download the "body" of the document based on the rights / access modes attributed to the document itself.
5. Authentication:
  - a. A rights management service / access mode to a service (see insertion) or a single document by a generic user. Authentication can be done in two different ways:
    - i. User Authentication: The system asks an external "LDAP server", dedicated to the specific institution that provides its user recognition
    - ii. IP Address Authentication: The system detects the IP address of the user machine that is interacting with a particular service and verifies it with the IP address-enabled IP address table
6. Access to the "body" of the document:

The "body" access / access service of a document verifies that each request conforms to the established access policy for all institutions and the right of access established for the Single document.

There are currently three types of access law:

  - a. Right of Access:
    - i. Free - the "body" of the document is made available to everyone
    - ii. Reserved - the "body" of the document is made available only to users authenticated via LDAP or IP address
    - iii. Protected - The body of the document is made available only to users authenticated via LDAP but are also the author / co-author of the document

### **The MySql database**

The relational database (MySql) is used to create indexes across a range of fields that make up the "bibliographic description" of the document to speed up search and list operations. It also keeps information about collections, institutions and their administrators "username and password", as well as information about assisted assignment of "document code".

A specific document describes the schema of the database and the individual tables that compose it.