Publications Office

Technical Annex IV

Overview of the system architecture of the DOI back office for the Publications Office

Update: 10 May 2018

Introduction

This document gives an overview of the technical infrastructure which enables Publications Office staff to assure the creation, the registration, the maintenance and the resolution of DOI names.

1. OP DOI RA functional architecture

The overall OP DOI RA system consists of several components and modules that enable Publications Office staff to assure the creation, the registration, the maintenance and the resolution of DOI names. Following, main components and modules are grouped and described according to their functional role:

- User-facing Web applications: components responsible for the collection of DOI metadata (Web editors, xml upload, web service), their preparation, transformation, submission to the registration workflow (web service), monitoring of the registation process (statistics, history) and the user administration (IANUS). Receive data from users and interact with:
    - DOI Registration module
- DOI Registration module (DOI Manager): is the component that contains the logic to process requests for DOI Registration and is responsible for metadata storage and DOI minting. Receives data from the User facing we application for the collection of DOI metadata and interacts with:
    - Administration Database (IANUS) to check for users' registration rights and to add accounting records
    - Metadata Database to create and update DOI metadata records
    - Local Handle Server (LHS) to create and modifyf Handle records (DOI).
- Crossref and DataCite deposit handler (Engine): is the component that processes the record, applies any transformation necessary to output a metadata record in the format required and sends it to the appropriate API for deposit on external systems. Receives data from the DOI manager and interacts with:
    - Crossref system
    - Datacite system.
- Authentication: in order to access all the system components, users must be authenticated by user-id and password, via Shibboleth for the b2c interactions and via HTTP basic authentication for web services and b2b interactions (see below). This allows the Publications Office and only the Publications Office to use the Registration Agency services.

All software and data elaboration is based on principles of platform independence. The metadata collection database is based on UTF-8, as well as data stored in the Local Handle Server database. Metadata transformation processes are developed in Java language. This allows maximum portability of the code base and native Unicode support.

The diagram below describes OP DOI RA functional architecture in production environment. A development environment, fully aligned with the production one, is available for testing.

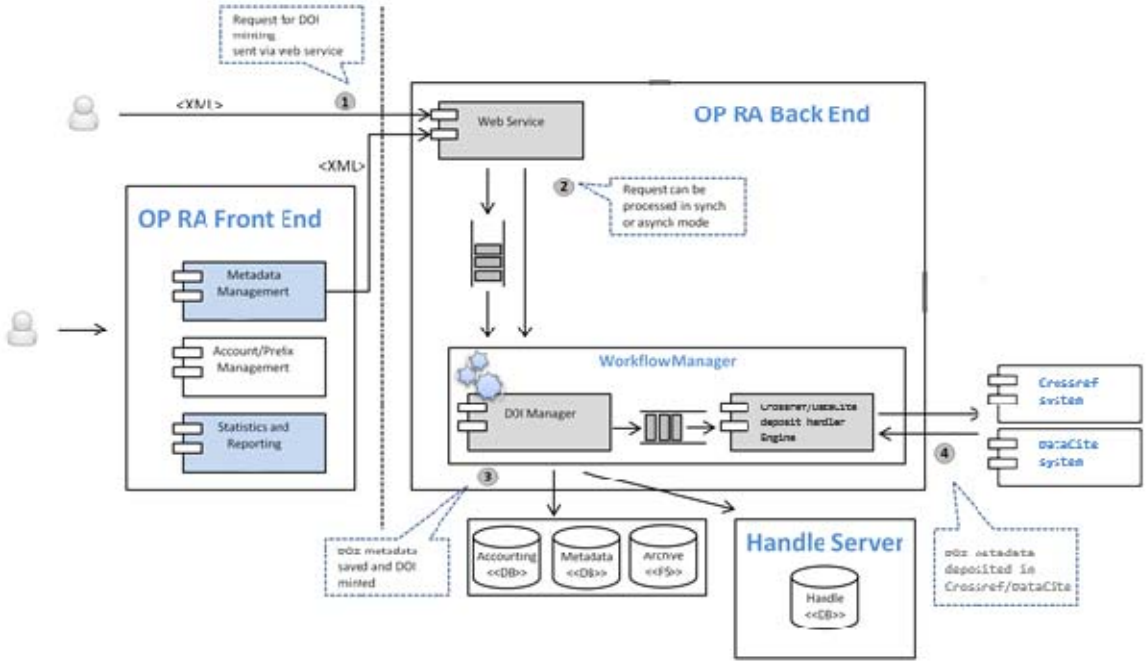

Figure 1 –OP DOI Registration Agency functional architecture (Production)

All components and modules have been developed for OP DOI RA, with the exclusion of the Handle System and IANUS suite that therefore are briefly described below in their specific use for OP DOI RA.

1.1 Handle System

The Handle System is a general-purpose global name service that allows secured name resolution and administration over the public Internet. The Handle System includes an open set of protocols that enable a distributed computer system to store DOI names of digital resources and resolve them into the information necessary to locate, access, and make use of the resources. The Handle System consists of Global root servers, Local Handle Servers, clients, and Proxy Servers.

OP DOI RA runs the Handle System package version 8.1.1. OP DOI RA runs its own Local Handle Server (LHS) and DOI Proxy. The DOI Proxy is the interface between the LHS and the HTTP protocol, thus allowing users to resolve DOIs using a web browser.

OP DOI Handle Server (administration) is de-coupled from OP DOI Proxy (resolution).
OP DOI Proxy interface (web page) is customised for OP and can be further customised. This means that the users can resolve DOIs in the Internet through a Publications Office “branded” server, for example at “dx.publications.europa.eu” or https://data.europa.eu/doi/.

1.2 Administration Application: IANUS

The IANUS suite is an infrastructural technology especially developed to control accesses to Internet-based services. IANUS technology is based on open system concepts (the independence from the DBMS platform, for example) and, together with its software components it is fully portable to Unix systems.
In the context of OP DOI RA, IANUS technology usage is limited to the management of access permissions, the actual access control being managed by Shibboleth identity provider.
In order to administer access to back office web services, IANUS technology is deployed.

Basic IANUS Suite's components are:

- IANUS Server: it is an extended Apache server, based on open-source components.
- Change Password: this facility enables registered users to change their password.
- Administration: it is a web tool to manage accounts for users and groups of users and to define rules for controlling users access. The main functionalities are:
    - User Maintenance: This function allows a system administrator (in this case, the Publications Office's DOI Registration Agency staff), to create and maintain "User" details, providing user-id and password for each user. Details of their name, address and contact details can also be maintained. Users can be allocated to one or more User-Groups.
    - User-Group Maintenance: This function provides the ability for a DOI Registration Agency administrator to define and maintain a "Group of Users", representing a User's organisation; two sets of name, address and contact details can be stored for each Group of Users, representing their commercial-contact and invoicing-contacts, VAT code/status and contract identifier.

The basic IANUS Suite has been extended with functionalities to manage DOI prefixes and their relations to Users and Groups of Users.

2. Hardware and software infrastructure

The OP Registration Agency (OP DOI RA) system components used to provide the Back Office services (see Technical Infrastructure) rely on hardware and software infrastructure composed of an application-presentation tier and a database tier. At each level, several infrastructural components are deployed. The deployment diagram below describes OP DOI RA infrastructure in production environment. A development environment, fully aligned with the production one, is available for testing.

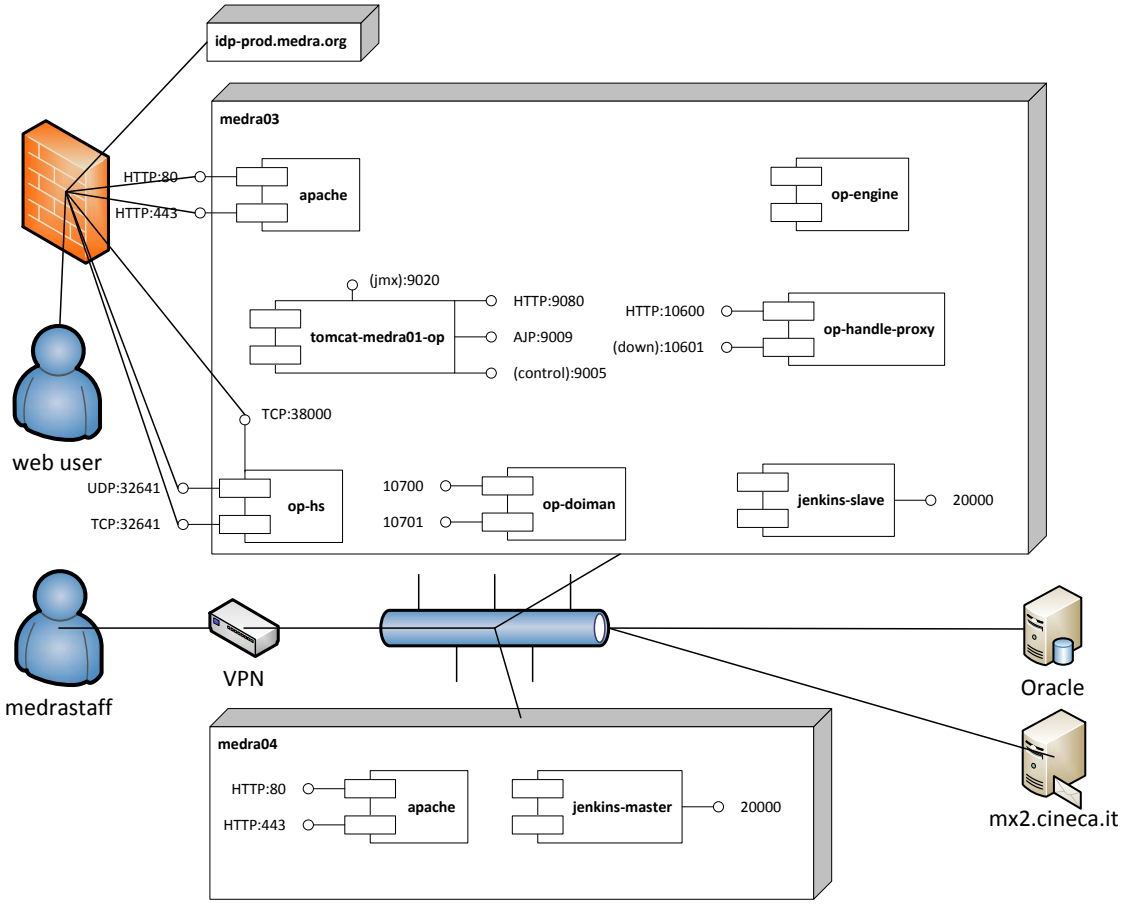

Figure 2 –OP DOI Registration Agency deployment diagram (Production)

The following sections detail the hardware components and the software components deployed at application-presentation tier and database tier.

2.1 Application-presentation tier

2.1.1 Software components:

- An application server (Apache)
- A servlet container (Tomcat)
- A continuous integration module (Jenkins)
- A static website
- A dinamic website (Java Web App)
- The DOI registration module (Java application, DOI manager)
- The Crossref and Datacite deposite handler (Engine)
- An identity provider for user authentication and authorization (Shibboleth IDP)
- The Local Handle Server (LHS)
- The Handle Proxy

2.1.2 Hardware components (server):

One Linux Virtual Machine based on:

- 2 Virtual CPU, 2.4 Ghz
- 32 Gbyte RAM
- 250GB disk space
- O.S.: Debian GNU/Linux 8.9 (jessie)
- Puppet provisioning software

2.2 Database tier

2.2.1 Software components

Components are supplied by Oracle database management system:

- LHS database
- Metadata collection database
- Administrative and accounting information database

2.2.2 Hardware components (server):

One stand-alone virtual machine running a dedicated Oracle Database 12c Standard Edition Release 12.1.0.2.0 - 64bit Production with the following characteristics:

- 2 Virtual CPU, 2.4 Ghz
- 16 GB RAM
- OS: Red Hat Enterprise Linux OS
- Oracle Tablespace: up to 500 GB

3. Hosting

3.1 Hosting infrastructure

The Publications Office Registration Agency services run on a virtual server, specifically designed for high performance and easy maintenance, high availability and convenient recovery.
The servers are currently located in an external data centre, located in Italy.
The data centre hosting OP DOI RA has a network infrastructure that allows to benefit of different types of connectivity:

- Local area network connectivity: redundant switches with independent power supply in order to ensure connectivity in case of failure of a single net component.
- Internet connectivity: the access to the Internet through the local area network can be provided with a band dedicated to the single server. The structure of the connectivity allows to have Internet links with different providers. By this feature, load balancing and automatic redirection of traffic is possible. This allows to preserve the connectivity in case of failure of one of the links. It is also possible to activate functionalities of Network Address Translation (NAT) and stateless Packet Filtering.
- Firewall and VPN services: it is possible to exploit advanced firewall functionalities (stateful packet filtering) to protect the client LAN against the most common attacks coming from the outside. This enables the possibility to have an endpoint of a Virtual Private Network (VPN). All net equipments (routers, switches, firewalls) are configured as clusters for high availability in order to guarantee connectivity also in case of hardware and software failures or maintenance activities.

OP DOI RA migration from "physical" machines, mounted in a blade center, to "virtual" machines avoids the possibility that the break of the internal switch in the blade center blocks the connectivity of the contained blades. Virtualization should also decrease the risk associated with blade break, because even in case of hardware problems, the virtual machine can be migrated to a working blade.

Technical details of the servers are provided in 2.1.2 and 2.2.2

3.2 Hosting service

The hosting service includes:

- Hardware, operating system, Oracle and software licenses
- Hardware and operating systems installation and management
- Oracle database installation, management and hosting
- Application software installation and management
- Systems and infrastructure tuning and customization
- Systems monitoring services (NetEye, Pingdom)
- Daily data saving and restoring (including incremental backup and disaster recovery)
- System assistance and network management from 9.00 to 18.00 CET on working days
- Security assessment and monitoring, including server and service monitoring

All necessary software licenses for components necessary for running the system have been acquired and included in the hosting services:

- Oracle 12c Standard Edition
- IANUS license.

Access to related technical documentation

Additional technical documentation can be found on the Publications Office Registration Agency website. For the purposes of this tender a guest account has been created. In order to access this documentation, please go to the website https://ra.publications.europa.eu. Guest access is provided using the following username and password:

Username: BKOFFICE_GUEST
Password: 06ZNGCRA