



WP8: JRA3 – Research on e-infrastructure for data and information management

D8.4

First Testbed Available

Expected date
M18



PROJECT DETAILS

PROJECT ACRONYM

NFFA-Europe

PROJECT TITLE

NANOSCIENCE FOUNDRIES AND FINE ANALYSIS - EUROPE

GRANT AGREEMENT NO:

654360

FUNDING SCHEME

RIA - Research and Innovation action

START DATE

01/09/2015

WP DETAILS

WORK PACKAGE ID

WP8

WORK PACKAGE TITLE

JRA3 – Research on e-infrastructure for data and information management

WORK PACKAGE LEADER

Stefano Cozzini (CNR-IOM)

DELIVERABLE DETAILS

DELIVERABLE ID

D8.4

DELIVERABLE TITLE

First Testbed Available

DELIVERABLE DESCRIPTION

This deliverable describes the implemented prototype and the tested procedures

EXPECTED DATE

M18 28/02/2017

ESTIMATED INDICATIVE PERSONMONTHS

6

AUTHOR(S)

Rossella Aversa (CNR-IOM), Stefano Cozzini (CNR-IOM)

PERSON RESPONSIBLE FOR THE DELIVERABLE

Stefano Cozzini (CNR-IOM)

NATURE

R - Report

DISSEMINATION LEVEL

- P – Public
- PP - Restricted to other programme participants & EC: (Specify)
- RE - Restricted to a group (Specify)
- CO - Confidential, only for members of the consortium

REPORT DETAILS

ACTUAL SUBMISSION DATE

NUMBER OF PAGES

20/02/2017

8

FOR MORE INFO PLEASE CONTACT

Rossella Aversa (CNR-IOM)

Tel. +39-040-3787515

Email: aversa@iom.cnr.it

Version	Date	Author(s)	Description / Reason for modification	Status
0	13/02/2017	Rossella Aversa	First draft	Draft
1	15/02/2017	Stefano Cozzini	Revision	Revision
2	17/02/2017	Thomas Jejkal	Revision	Revision
				Choose an item.
				Choose an item.
				Choose an item.
				Choose an item.



Contents

Contents	4
Executive Summary	5
1. Prototype's deployment	5
1.1 The IDRP	6
1.2 The NFFA portal	7
1.3 The Data Archives	7
2. Conclusions & Perspectives	7
References	8

Executive Summary

This deliverable reports a short description of the first NFFA Information and Data Repository Platform (IDRP) prototype developed by KIT [1], in collaboration with CNR-IOM [2] and Promoscience srl [3], and deployed on CNR-IOM OpenStack cloud infrastructure [4]. The prototype implements what is described in detail in D8.2 deliverable "Design of the finalized architecture" [5].

All the components of the prototype have been deployed as independent virtual machines. Each single element, and the interaction with the others will be briefly described in the next section.

A user guide describing the basic procedures to interact with the IDRP can be found at [6] and [9].

These web pages will be constantly updated and is intended to become the manual for the IDRP infrastructure usage.

1. Prototype's deployment

This section describes the components of the first NFFA prototype, based on the IDRP concept described in the NFFA proposal (Figure 1) and refined in deliverable D8.2.

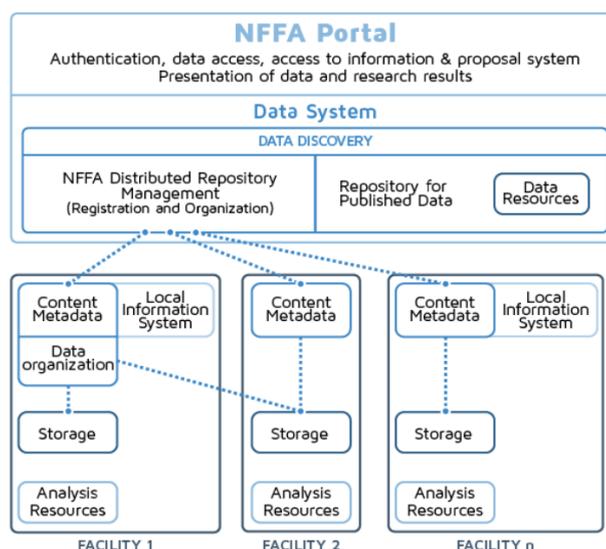


Figure 1: Conceptual architecture, as presented in the NFFA proposal.

The prototype installation has been deployed on the CNR-IOM OpenStack cloud infrastructure. As shown in Figure 2, it is composed of different elements installed on different virtual machines, representing the layers of the architecture. The IDRP is indicated in blue. The orange box represents the NFFA portal, while the grey ones represent external services, e.g. B2SHARE, an external service provided by the EUDAT [7] project to publish measurements, or data management resources located at NFFA facilities. Purple boxes represent instruments producing data, which is stored in local data archives. The data flow is represented by black arrows.

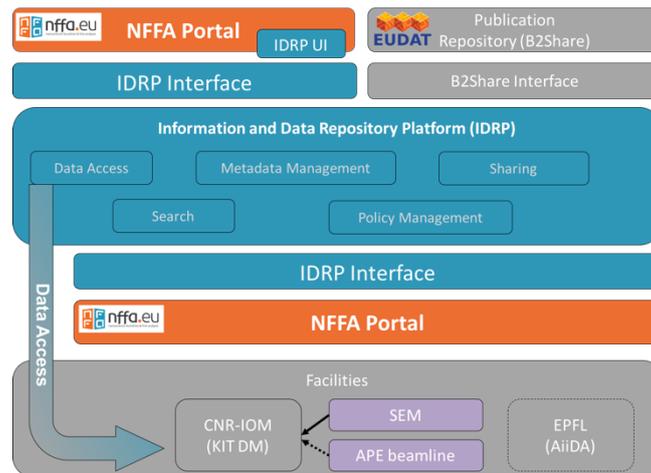


Figure 2: Schematic view illustrating the prototype installation and the interaction among the different components: the IDRP (blue), the NFFA portal (orange), the external service of B2SHARE (grey boxes at the top), the Data Archives at the local facilities (grey boxes at the bottom), and the instruments (purple boxes). All components are installed for testing at the CNR-IOM cloud infrastructure. Dashed lines indicate elements and data flows that have been already planned but not yet implemented or tested.

1.1 The IDRP

The IDRP is the core of the architecture, connected both to the data archive of the local facility and to the NFFA portal [8]. It has been installed on an Ubuntu virtual machine exposed at the public IP address 147.122.7.215, which for the time being is accessible only through a VPN.

The IDRP exposes several RESTful endpoints providing the following methods:

- to register, manage and retrieve metadata
- to register and retrieve to access and link metadata and data stored in the local repositories
- to manage the authorization of metadata and data access
- to interact with the NFFA portal in order to retrieve proposal and user information.

A documentation of all available endpoints can be found at [9].

The main purpose of the RESTful endpoints is to provide a defined interface to the data platform. On the one hand, this interface will be used in future to register data assets as soon as they have entered to local data archive. On the other hand, this interface allows to build graphical user interfaces for querying and visualizing data and metadata of NFFA proposals.

For prototype installation, a basic user interface is provided as part of the IDRP installation. This user interface allows scientists to register, modify and retrieve their metadata and data. Furthermore, it allows to share and publish measurements, either to IDRP users or worldwide using the EUDAT B2Share service. Furthermore, it is integrated seamlessly into the NFFA portal, which serves as single entry point to the information portal as well as to the IDRP.

1.2 The NFFA portal

The NFFA portal deployed in the prototype is a local installation on a Windows virtual machine exposed at the public IP address 147.122.7.226, which for the time being is accessible only through a VPN.

The NFFA portal exposes a REST interface to dialog with the IDRPs to provide authorization and authentication services, and information on the basic metadata related to the proposals submitted by NFFA users.

A number of fictitious proposals with status 'accepted' have been created on the NFFA portal for testing purposes.

1.3 The Data Archives

The NFFA landscape includes many facilities, in which different data management systems are available, e.g. KIT Data Manager [10], ICAT [11], iRods [12], NoMad [13], and AiiDA [14].

At the CNR-IOM, the data management system adopted is based on the KIT Data Manager. The KITDM@CNR instance has been installed on an Ubuntu virtual machine exposed at the public IP address 147.122.7.222, which for the time being is accessible only through a VPN.

Instruments data at the facility is ingested/downloaded to/from the KITDM@CNR, which creates the hierarchical metadata structure needed to store them.

In this prototype we created an ad-hoc plugin for the SEM instrument, with automatic metadata inclusion based on machine learning technique for image recognition. A considerable set of SEM images has been loaded on the KITDM@CNR, and some of them have been registered and 'published' on the IDRPs.

An additional plugin for the APE beamline has been already developed, but not yet fully implemented in the prototype.

At the EPFL [15], the data management system adopted is the AiiDA system, which stores simulation data. A link to the REST interface of the IDRPs has been designed and is currently under development.

2. Conclusions & Perspectives

In this document, the first NFFA IDRPs testbed developed at KIT, in collaboration with CNR-IOM and Promoscience srl, has been shortly presented. All the details about the architecture are reported in D8.2 [4], a description of current functionalities can be found at [6] and [9].

An overview of the deployed installation has been given, together with details on the single components which have been tested.

A link to the user guide with an introduction on the basic procedures allowing to interact with the prototypical IDRPs user interface has been also provided.

The present prototype will be enhanced and completed at M30, and will be described in the D8.5 "Testbed fully deployed, including DASS services".

Next steps are testing and improving the prototype in terms of performance and usability. Afterwards, remote data asset registration using the IDRP RESTful endpoints will be shown exemplarily based on AiiDA and ICAT in order to automate data registration.

References

- [1] Karlsruhe Institute of Technology website: <https://www.kit.edu/english/>
- [2]: CNR-IOM website: <https://www.iom.cnr.it/>
- [3] Promoscience srl website, <http://www.promoscience.com/>
- [4] CNR Openstack cloud, <http://nimbo.escience-lab.org/dashboard/auth/login/>
- [5] Thomas Jejkal, "Design of the finalized repository architecture", <http://intranet.nffa.eu/DocumentRepository>
- [6] Thomas Jejkal, "NFFA Information and Data Repository Platform", <http://ipelsdf1.lsd.f.kit.edu/nffa/idrp/manual/index.html>
- [7] EUDAT website, <https://www.eudat.eu/>
- [8] NFFA portal, <http://nffa.eu/>
- [9] Thomas Jejkal, "Information and Data Repository Platform - RESTful API", <http://ipelsdf1.lsd.f.kit.edu/nffa/idrp/api/index.html>
- [10] Karlsruhe Institute of Technology, KIT Data Manager, <http://datamanager.kit.edu/index.php/kit-data-manager>
- [11] ICAT project website, <https://icatproject.org/>
- [12] iRods website, <https://irods.org/>
- [13] NOMAD website, <http://repository.nomad-coe.eu/cms/>
- [14] AiiDA website, <http://www.aiida.net/>
- [15] EPFL website, <https://www.epfl.ch/index.en.html>