Deliverable D6.2

| | | |
|---|---|---|
| Project Title: | World-wide E-infrastructure for structural biology | |
| Project Acronym: | West-Life | |
| Grant agreement no.: | 675858 | |
| | | |
| Deliverable title: | Report on Repository implementation | |
| WP No. | 6 | |
| Lead Beneficiary: | 1: STFC | |
| WP Title | Demonstrator | |
| Contractual delivery date: | 30 November 2017 | |
| Actual delivery date: | 30 April 2018 | |
| WP leader: | Tomas Kulhanek | STFC |
| Contributing partners: | STFC, CIRMMP, INSTRUCT, MUNI | |

Deliverable written by Tomas Kulhanek, Fiona Sanderson, Andrea Giachetti

# Contents

# 1 Executive summary

Large facilities already have data management services, for example ISPYB at several European synchrotrons. However it at smaller facilities it is all too common that the user simply returns home with a USB stick.

This deliverable developed a Repository for experimental data, which is installable locally or on cloud provision.

It covers the following phases of the structural biology data life cycle:

- Data acquisition
- Data preprocessing
- Publishing to long term preservation

This report describes the first release of this software. It was integrated with Instruct's ARIA visit management system, in collaboration with partner Instruct, in order to provide a seamless online experience for structural biologists.

# 2 Project objectives

With this deliverable, the project has reached or the deliverable has contributed to the following objectives:

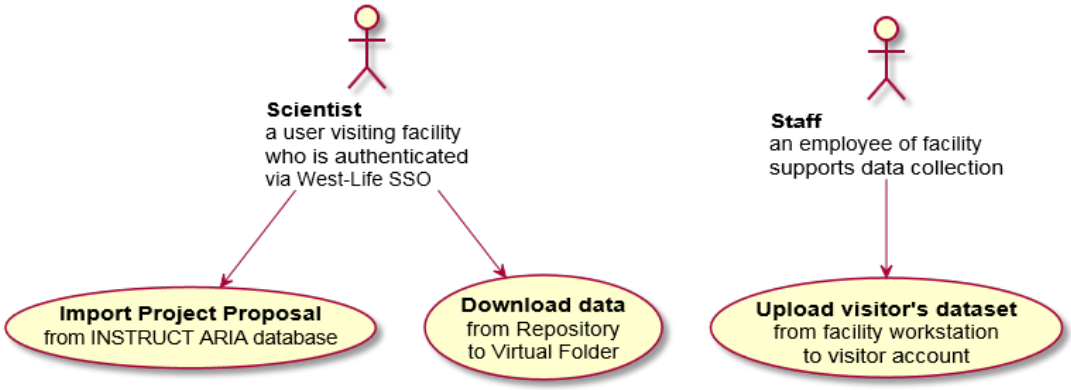
| No. | Objective | Yes | No |
|---|---|---|---|
| 1 | Provide analysis solutions for the different Structural Biology approaches | X | |
| 2 | Provide automated pipelines to handle multi-technique datasets in an integrative manner | | X |
| 3 | Provide integrated data management for single and multi-technique projects, based on existing e-infrastructure | X | |
| 4 | Foster best practices, collaboration and training of end users | X | |

# 3 Detailed report on the deliverable

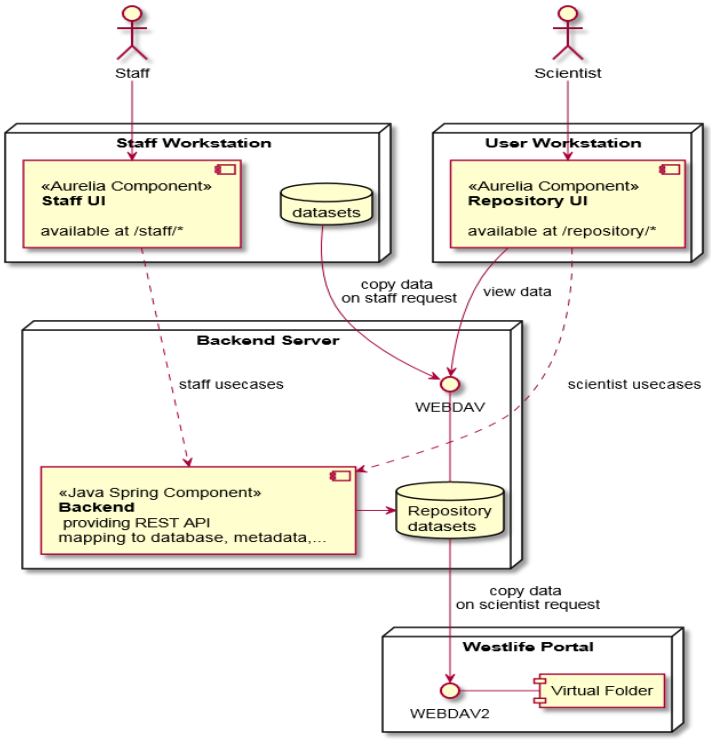
## *3.1 Overall architecture design*

The repository implementation covers the following main use cases:

- Project proposal import from INSTRUCT ARIA database.
- Dataset upload from facility workstation to visitor account.
- Dataset download from Repository to Virtual Folder.

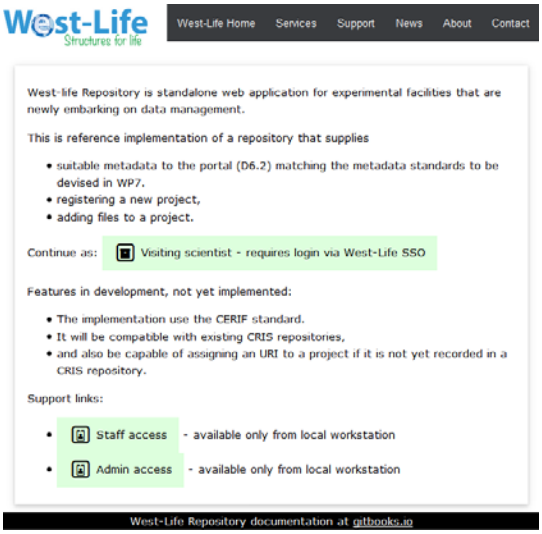


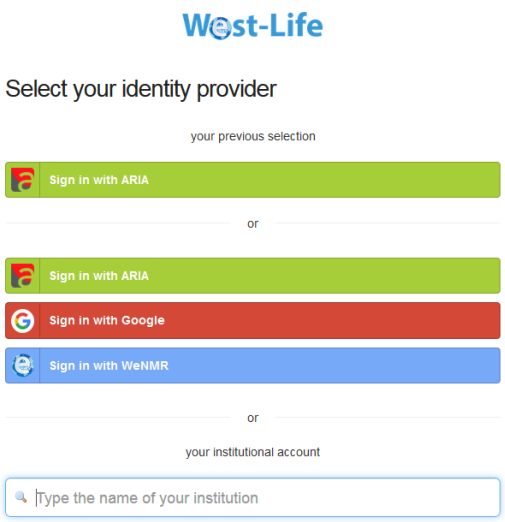

The Repository architecture is designed with REST web services provided by the backend (implemented in Java Spring framework) and REST client and web UI provided by frontend (implemented in Javascript (ECMAScript6) and Aurelia JS framework) distinguishing between identified stakeholders (Visiting Scientist and Staff):



## 3.2 Authentication

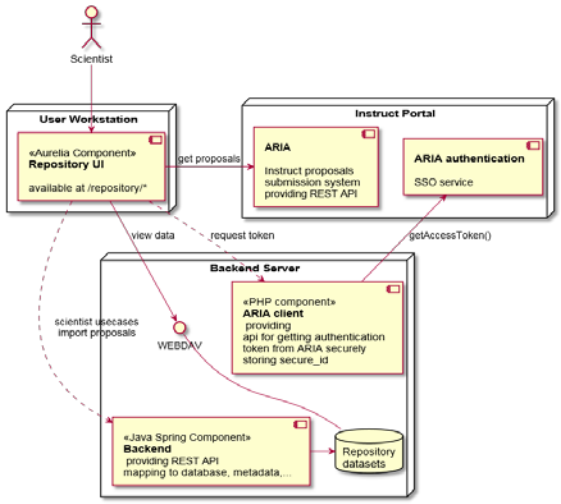
Authentication is done by West-Life SSO. The Repository is a common service provider, which needs to be registered within West-Life SSO identity provider based on SAML. The mod_auth_mellon module is used to secure selected context path. Unauthenticated requests are redirected to West-Life SSO login site. After login, it is redirected back with appropriate attributes set for underlying services. Additional http headers are set for services behind apache http server in order to identify the user.




## 3.3 Project proposal import

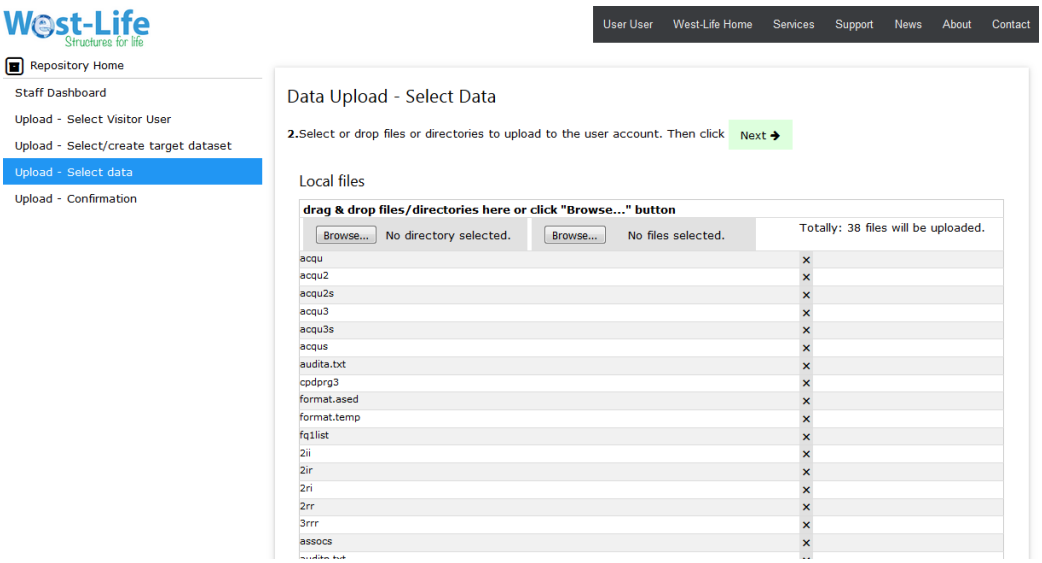
A user authenticated via West-Life SSO can see their private projects and datasets stored within the repository implementation. However, project proposals need to be imported from the Instruct database using the ARIA API first. The ARIA OAuth client components are implemented as PHP scripts integrated in backend components. The authentication token is then used by the Javascript frontend to perform AJAX calls.




Detailed ARIA integration and amendments to ARIA specs: [https://h2020-westlife-eu.gitbook.io/virtual-folder-docs/repository/developers-guide/aria-integration](https://h2020-westlife-eu.gitbook.io/virtual-folder-docs/repository/developers-guide/aria-integration)

## 3.4 Dataset upload

A staff user (authenticated with backend, not with SSO) can upload data acquired on a device from the workstation on behalf of a visitor scientist.

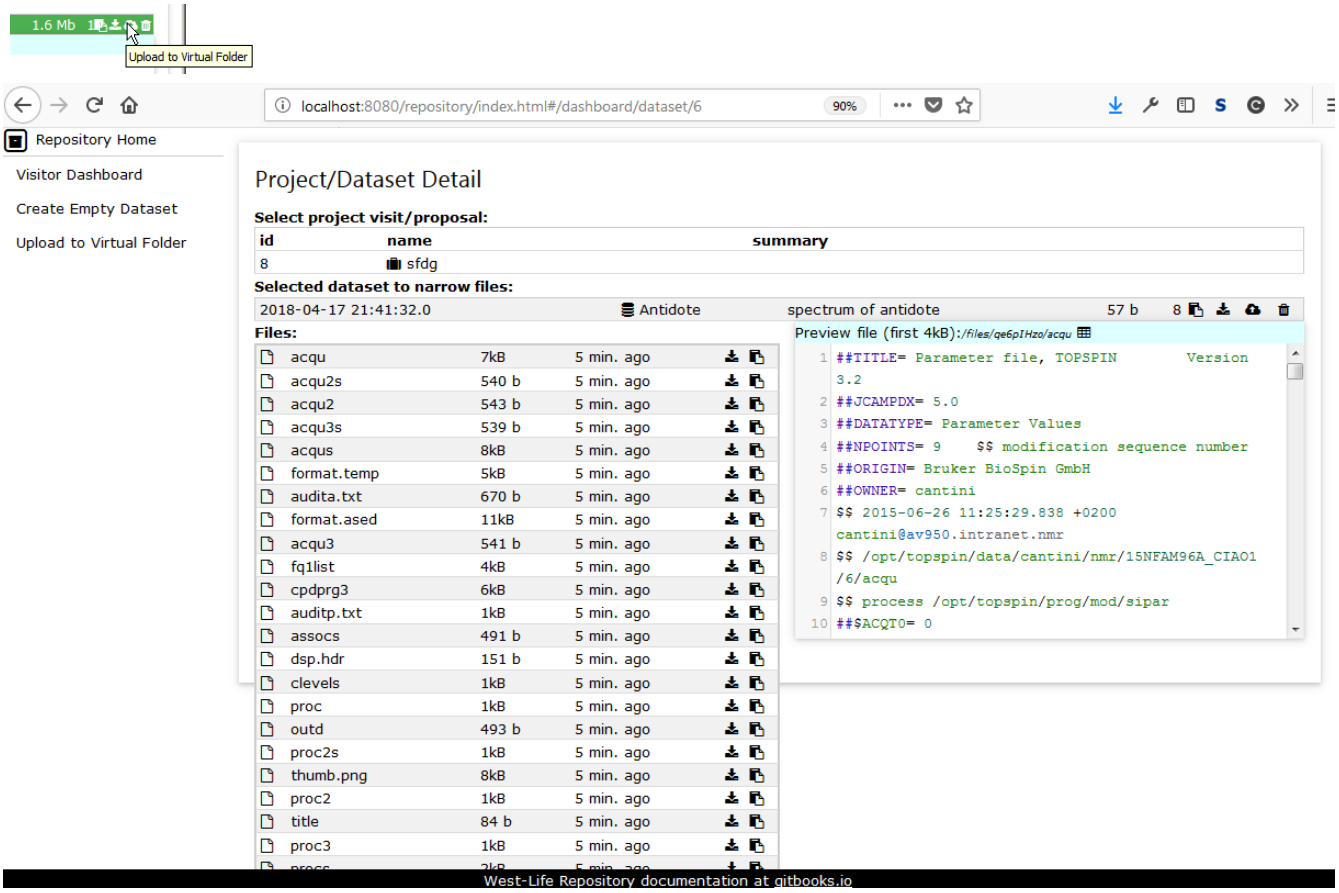



## 3.5 Backend and persistence configuration

The backend is based on the Spring MVC java framework. The local administrator is registered at the deployment of the service, using the Spring security API. The persistence is managed by Spring with JPA, using Hibernate as persistence provider. The following schema is implemented in order to manage accounting information and the association of local files to projects and datasets.

## 3.6 Dataset Download to Virtual Folder

A dataset files uploaded by staff user can be stored within Repository for limited time – as the amount of data can be enormous – facility cannot guarantee the long term preservation of data. Therefore, a feature to download data from Repository to Virtual Folder is available for Visitor Scientist.




## 3.7 Standards

The dataset upload is done via WEBDAV api provided by server endpoint referring to visitor storage. Dataset Download to Virtual Folder is done via WEBDAV api, WEBDAV client capabilities are implemented within server service.

Web interface file manager can browse the content of user storage and shows first 2 kB of the selected data file.

## 3.8 Installation and documentation

Installation is made as script configuring the server into the web server with backend and fronted providing endpoints and web UI. It is based on Scientific Linux 7.x, Centos 7.x or any derivative of RHEL 7.x. Vagrant script is provided for local test deployment. Installation instructions are documented at https://h2020-westlife-eu.gitbooks.io/virtual-folder-docs/content/repository/

# References cited

Repository documentation https://h2020-westlife-eu.gitbooks.io/virtual-folder-docs/content/repository/
Repository source codes https://github.com/h2020-westlife-eu/wp6-repository