Deliverable D5.3

Project Title:	World-wide E-infrastructure for structural biology	
Project Acronym:	West-Life	
Grant agreement no.:	**675858**	
Deliverable title:	Prototype of the new VRE portal functionality	
Lead Beneficiary:	LUNA	
Contractual delivery date:	Month 6	
Actual delivery date:	Month 6	
WP No.	5	
WP Title	Support	
WP leader:	Alexandre Bonvin	Partner: Utrecht University
Contributing partners:	Luna	

Deliverable written by Francois Ruty, Luna

1 Executive summary

The purpose of this deliverable is to set up a prototype of the VRE (Virtual Research Environment) within the project portal previously established in D5.1. The first version of the VRE is accessible at portal.west-life.eu (in the first instance, access is provided on demand).

The overarching goal of the VRE is to connect the many West-Life e-science portals to one another, as well as with academic storage repositories. Scientists must be able to use various WestLife e-science portals in an integrated manner, that is to say have a central location to manage all their data analysis and modelling. They must have the technical means to shuffle data between portals in an easy way, without for example having to transfer files via their own desktop over low-speed internet connections. Additionally, scientists shall have easy access to cloud resources, to scale and speed up cpu-intensive analyses or simulations. Thus, the aim is to provide a glue between existing academic services: West-Life e-science portals, academic storage repositories (e.g. EUDAT), and grid/cloud resources (e.g. EGI).

Three core features of the VRE are:

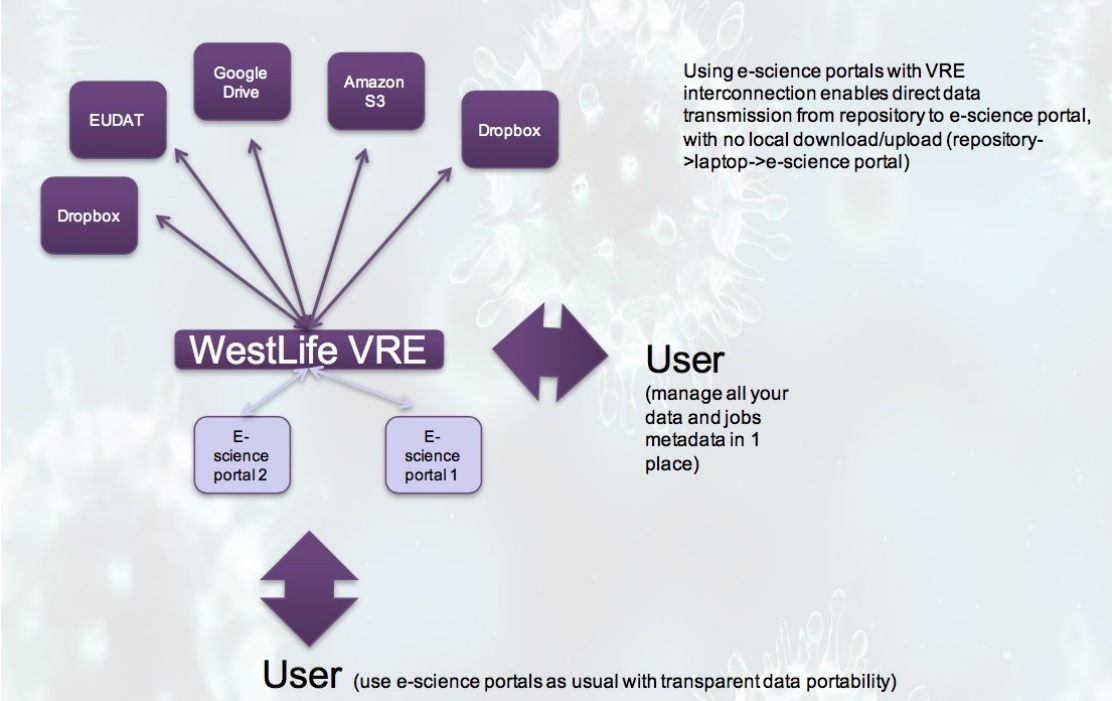
1. An aggregated view of all the scientist's data across personal and central repositories (EUDAT, Dropbox etc)
2. The ability to send job results directly from e-science portals to data repositories (no need to download/upload to a local desktop)
3. Storage of the metadata for all jobs submitted via the portal

These features are realised through specific data connections:

- The WestLife VRE connects to existing APIs of data repositories
- E-science portals connect to the WestLife VRE, using a simple procedure that minimises the disruption to existing services

The WestLife VRE shall provide:

- “Data location awareness” to e-science portals (where to fetch and dump data)
- Jobs metadata storage for e-science portals
- Aggregated jobs overview for end users
- Aggregated data overview for end users



This report describes a prototype for the West-Life VRE which addresses the above features. There is a particular focus on interconnection with academic storage repositories.

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

The VRE prototype is currently focused on interconnection with academic data storage repositories. Closer interconnection with the e-science portals provided by West-Life partners is the next step. We chose to address academic storage interconnection first since most of them come with an existing public API, while closer integration with the e-science portals will require small updates from portal operators.

Status

The current VRE prototype can connect to the EUDAT API, Dropbox, Google Drive and Amazon S3, and enables end users to have an aggregated view of all their data across all repositories (see screenshots in the Appendix).

It is also possible to add further non-commercial storage. The current VRE prototype can connect to any API, and so West-Life consortium members can set up an object store (with APIs) on the hardware they contribute to the project, and the VRE can include those storage services in the aggregated view it provides.

Users have a settings section on the VRE where they can allow the VRE access to their academic repositories.

Users can log on the web portal and view a unified folder structure that sums up what they have in all the repositories that they have connected to their VRE account. They can upload and download files, and create folders in any of their repositories, from the VRE. They can create shareable links to share files with other people.

Hardware requirements

The current VRE runs on Ubuntu 14.04 and does not yet store data locally – future versions will store metadata. The hardware requirements to run the web portal are low: a few 100s MB of disk space, less than 1GB of RAM and 1 CPU thread are necessary.

The lightweight nature of the West-Life VRE portal is a key requirement to make sure running costs are as low as possible, to ensure deployment is easy, and to ensure sustainability beyond the end of the project.

Software Technology

For the web portal, we use Python/Django for server code, and AngularJS for JavaScript code. It connects to the EUDAT B2DROP service by the standard WebDAV interface, and it uses the suppliers' custom APIs to connect to DropBox, Google Drive, and Amazon S3. An abstraction layer unifies the four into a single API for access to data repositories.

Other VRE Features

Some work was carried out to design the architecture of code snippets that West-Life e-science portals operators will be asked to include in their portals, in order to interface with the VRE. This has been circulated to other partners for comments.Those code snippets will allow e-science portals to query the VRE and fetch data locations. The VRE will thus abstract away the various storage repositories and will provide a single interface to e-science portals.

We expect the VRE to provide this unified interface to testbed WestLife portals within 3 months.

Note: all source code (API code snippet, web application etc) will be publicly available to ensure that the WestLife portal lasts beyond the end of the H2020 project.

Interaction with other WPs

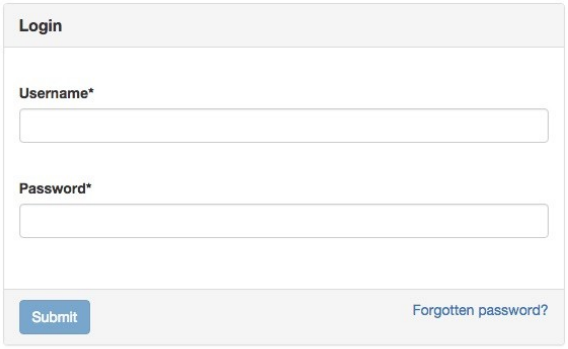
The work carried out here on an aggregated view of multiple data repositories is tightly coupled with deliverable D6.1 on providing the user with a virtual folder of distributed data. Work is ongoing to make sure the two efforts complement each other well.

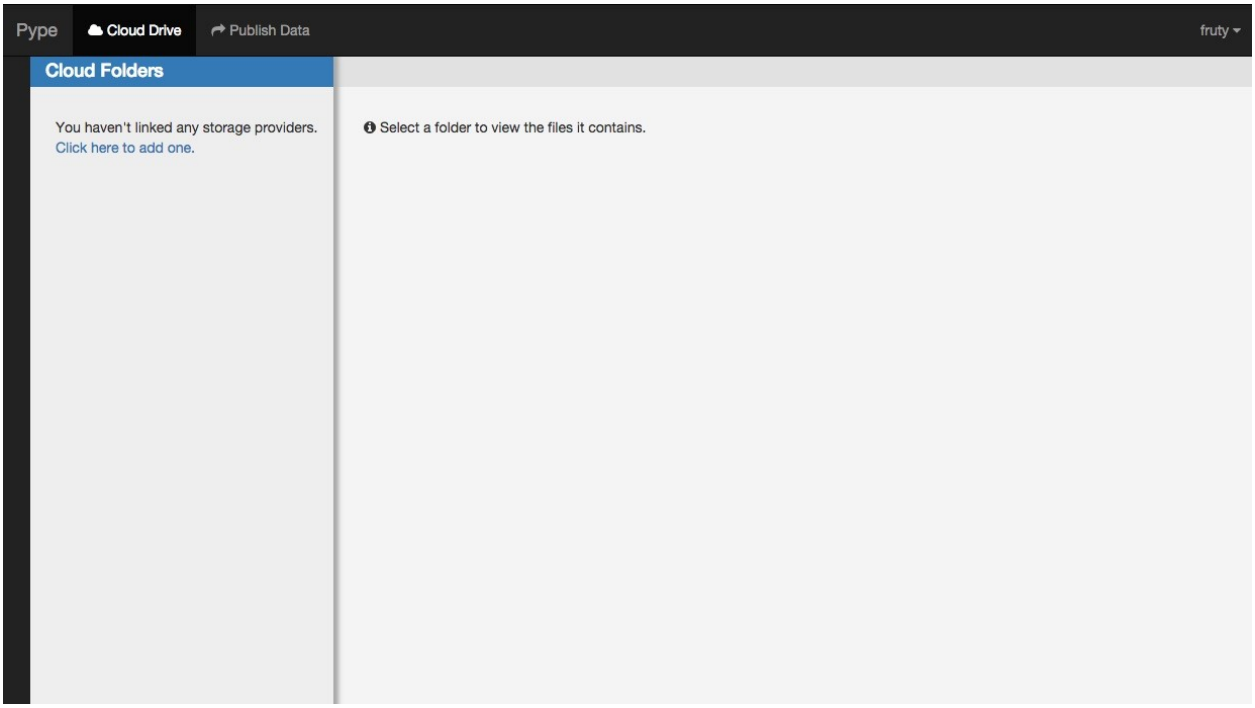
There are also on-going discussions with WP4 to make sure the small upgrades required from portal operators, for interconnection with the VRE, do not conflict with the backend job submission architecture evolution.

Appendix 1: Screen shots

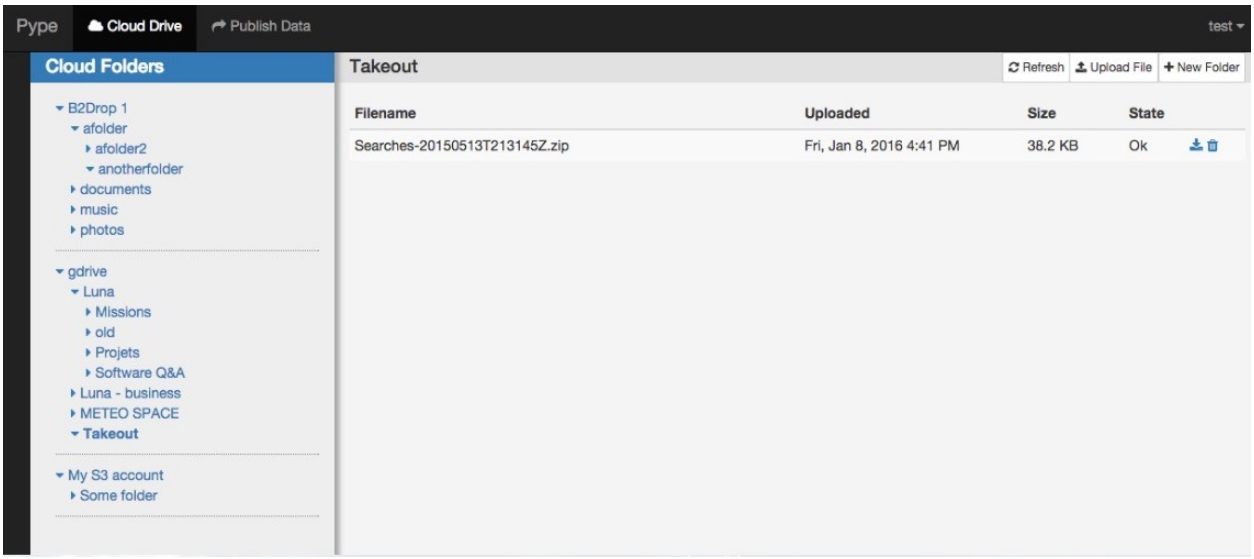
Data Management

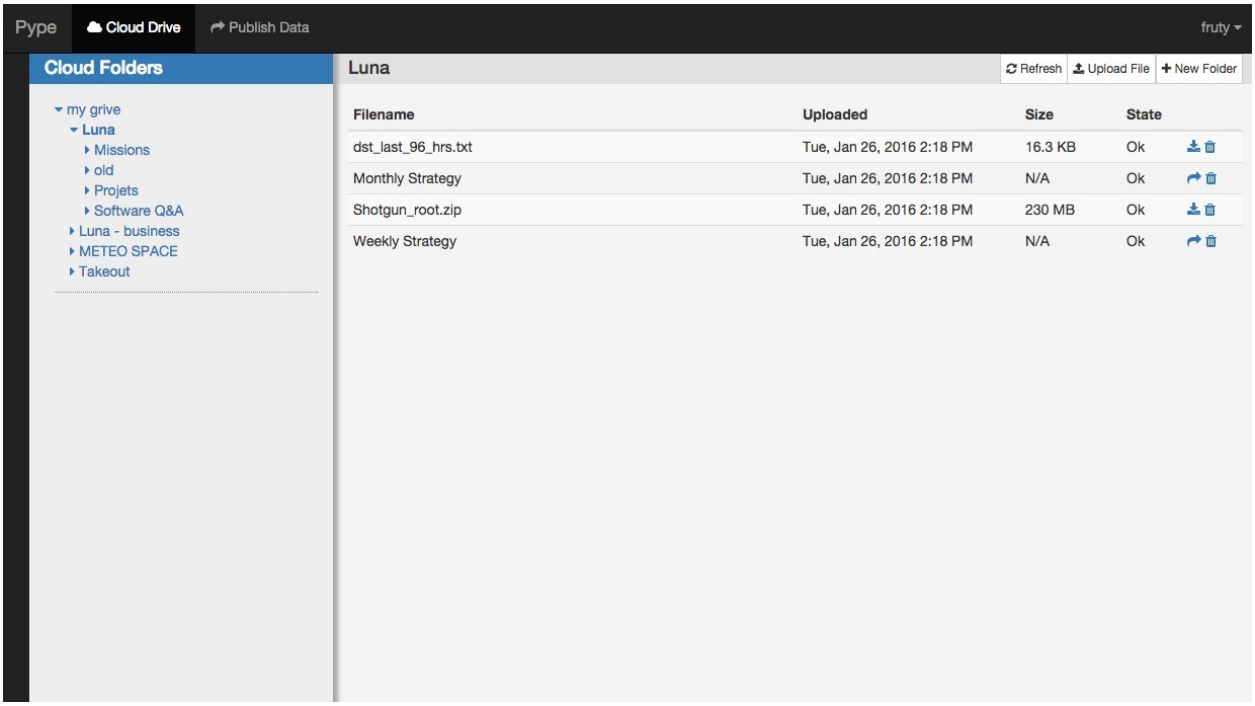
What is visible for partners: (after private login url)





On the screenshot above you can link your WestLife VRE account with your storage repositories (EUDAT, Google Drive etc….). The linking procedure is adapted to each provider. For instance, if you attempt to link a Google Drive account, it will open a new page where you have to login on your google account and allow the VRE access to your drive.

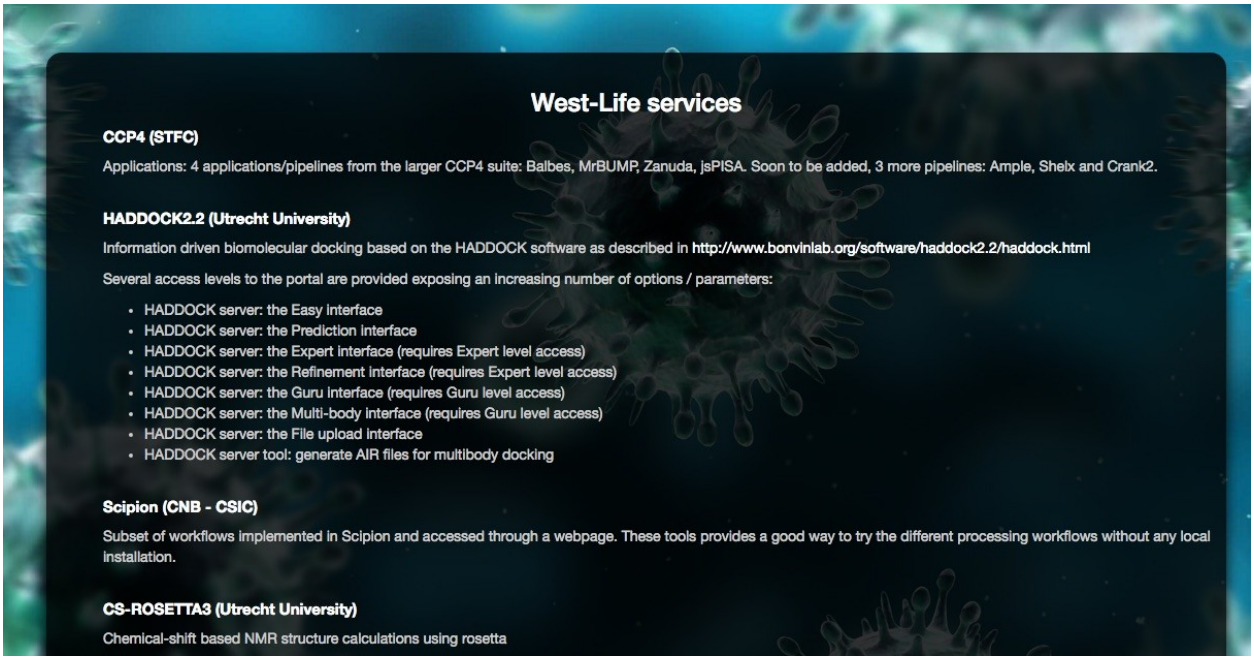

Pype
Cloud Drive
Publish Data
test
Cloud Folders
B2Drop 1
afolder
afolder2
anotherfolder
documents
music
photos
gdrive
Luna
Missions
old
Projets
Software Q&A
Luna - business
METEO SPACE
Takeout
My S3 account
Some folder
Takeout
Refresh
Upload File
New Folder
Filename
Uploaded
Size
State
Searches-20150513T213145Z.zip
Fri, Jan 8, 2016 4:41 PM
38.2 KB
Ok


Pype
Cloud Drive
Publish Data
fruty
Cloud Folders
my grive
Luna
Missions
old
Projets
Software Q&A
Luna - business
METEO SPACE
Takeout
Luna
Refresh
Upload File
New Folder
Filename
Uploaded
Size
State
dst_last_96_hrs.txt
Tue, Jan 26, 2016 2:18 PM
16.3 KB
Ok
Monthly Strategy
Tue, Jan 26, 2016 2:18 PM
N/A
Ok
Shotgun_root.zip
Tue, Jan 26, 2016 2:18 PM
230 MB
Ok
Weekly Strategy
Tue, Jan 26, 2016 2:18 PM
N/A
Ok

On the 2 screenshots above, you can see an example of an aggregated view with Google Drive, EUDAT B2DROP and Amazon S3 accounts.Academic partners can set up further object storage. Most of those systems expose an API very similar to Amazon S3 so integration of custom academic storage services is easy.

Job Submit

WestLife portals are operational from day 1



WestLife portals are already operational and available from the West-Life portal, but they are not yet integrated into the VRE. The VRE currently only provide an aggregated view of all data. In the near future, users will go to their usual e-science portals, and instead of having to upload data on those web interfaces, they will be able to select data from the VRE aggregated view, directly in the e-science portals interfaces.