Deliverable D5.5

Project Title:	World-wide E-infrastructure for structural biology
Project Acronym:	West-Life
Grant agreement no.:	**675858**
Deliverable title:	D5.5 VRE-integrated PDBe Search and Query API
WP No.	5
Lead Beneficiary:	EMBL
WP Title	Support
Contractual delivery date:	M18
Actual delivery date:	M18
WP leader:	Alexandre Bonvin	Utrecht University
Contributing partners:	EMBL-EBI, STFC

Deliverable written by Nurul Nadzirin

Contents

1 Executive summary

We have successfully integrated PDBe search and query REST API's in the VRE portal. This was done via the integration of several PDBe web components and the autocomplete search functionality within the Virtual Folder. Nine web components were integrated, namely PDB Links, PDB Prints, PDB Topology Viewer, Sequence Feature Viewer, PDB UniProt Viewer, PDB_REDO, PDB Residue Interactions, PDB 3D Complex, and LiteMol. These components utilize the query API's, which satisfies the requirement for the deliverable. The autocomplete functionality has also been integrated, which uses the PDBe search API thus also fulfilling the deliverable.

Demonstrations for the components and the autocomplete feature can be found on the portal under the Virtual Folder section (https://portal.west-life.eu/virtualfolder/).

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   Integration of PDBe query API

Several PDB components have been integrated to the Virtual Folder, which satisfy the deliverable of the integration of PDBe query API. Currently, the components can be accessed via either a tab found in the Virtual Folder's File Manager (https://portal.west-life.eu/virtualfolder/filemanager.html) or via a stand-alone demo called the 'Dataset demo' (https://portal.west-life.eu/virtualfolder/test/index-dataset.htmlhttps://portal.west-life.eu/virtualfolder/index-dataset.html).
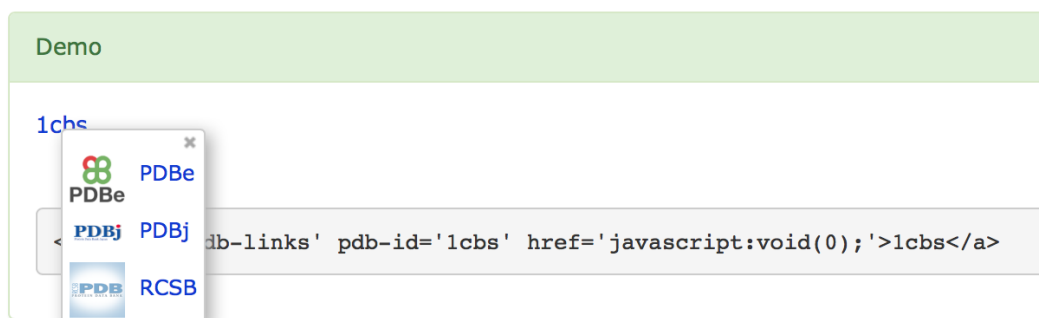
The web components are:

- **PDB Links**

Figure 1: A demo of the PDB Links component

> The PDB Links component adds a popup containing links to all the wwPDB resources for a given PDB entry. The popup is displayed when user clicks on a hyperlink to which the component is applied.
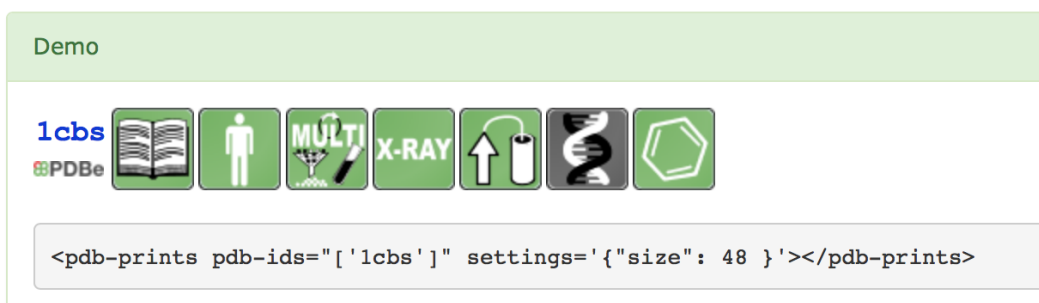
- **PDB Prints**

Figure 2: A demo of the PDB Prints component

> A PDBprint for a PDB entry is a collection of PDB logos displayed in a specific order, where each icon represents a well-defined category of information.

> In PDBprints the following categories are included:
> o   Primary citation: has the PDB entry been published?

West-Life Deliverable D

- Taxonomy: what is the source organism of the biomacromolecule(s) in the entry?
- Sample-production technique: how was the sample of the biomacromolecule(s) obtained?
- Structure-determination method: which experimental technique(s) was used to determine the structure and was the experimental data deposited?
- Protein content: does the entry contain any protein molecules?
- Nucleic acid content: does the entry contain any nucleic acid molecules (DNA, RNA or a hybrid)?
- Heterogen content: does the entry contain any ligands (such as inhibitors, cofactors, ions, metals, etc.)?

- **PDB Topology Viewer**

The topology viewer depicts the secondary structure of a protein in a 2D representation, taking into account the interactions of these secondary structure elements. This leads to a consistent display of sheets and domains in the structure. For PDB entries with multiple copies of a protein, the best chain is used. The topology viewer also depicts value-added annotation from SIFTS including residue-level mapping to UniProt, sequence families (Pfam), structure domains (SCOP, CATH) and structure quality from wwPDB validation reports.
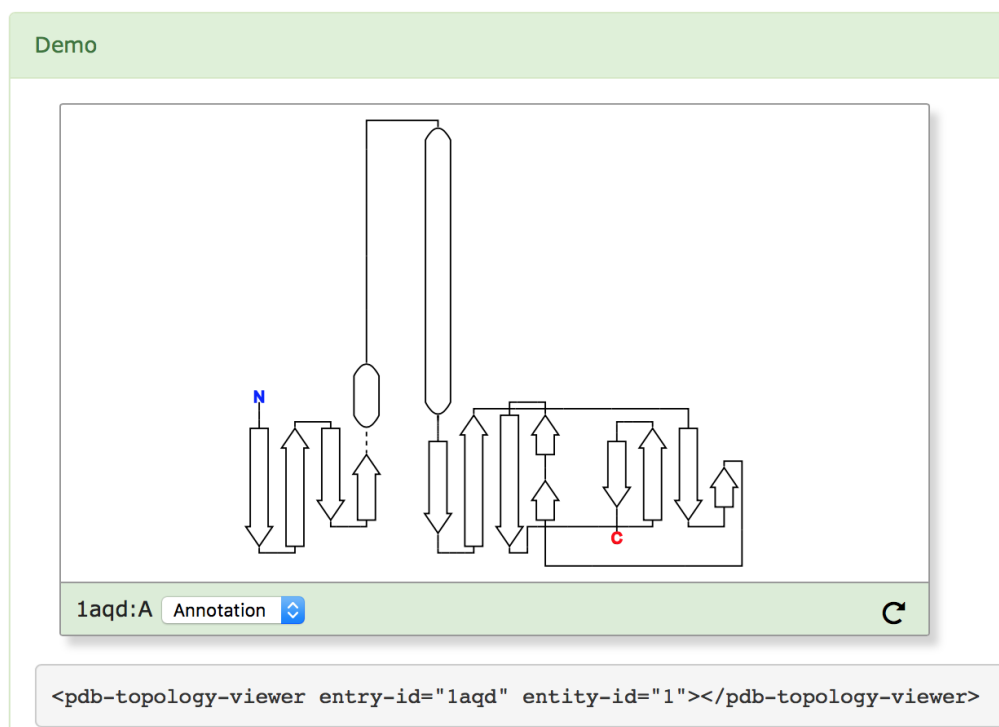


```
<pdb-topology-viewer entry-id="1aqd" entity-id="1"></pdb-topology-viewer>
```

Figure 3: A demo of the PDB Topology Viewer component

- **Sequence feature View**

The sequence-feature view shows a linear representation of the sequence of the protein in a PDB entry and depicts value-added annotation from SIFTS. This including residue-level mapping to UniProt, sequence families (Pfam), structure domains (SCOP, CATH), mutations, binding-site residues, structure quality and secondary structure. By default the sequence-feature view shows the chain that has the maximum number of observed residues.
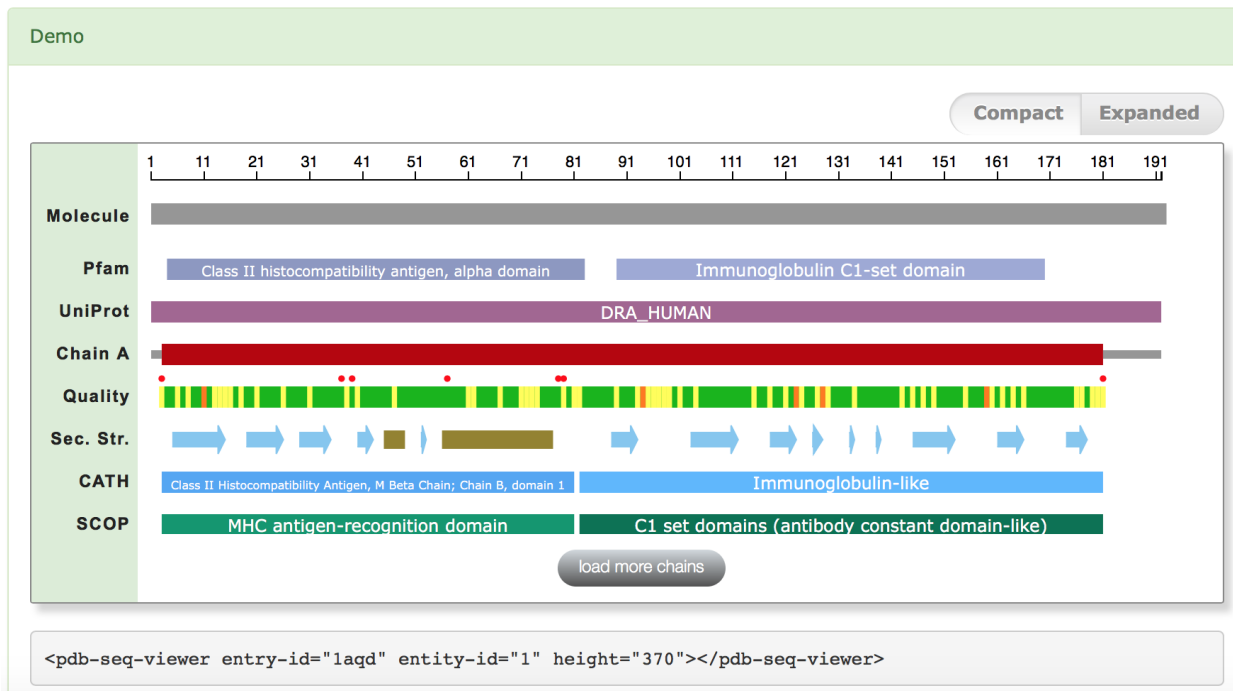
Figure 4: A demo of the Sequence Feature viewer component

- **PDB UniProt Viewer (UniPDB)**

The PDB UniProt viewer displays a summary of PDB entries contains a sequence mapped to a particular UniProt code. The mapping indicates what coverage of a UnipProt accession is available in the PDB archive. Graphics also indicate whether the sequence in a given PDB entry differs from that in Uniprot (for instance, it contains engineered mutations or expression tags).
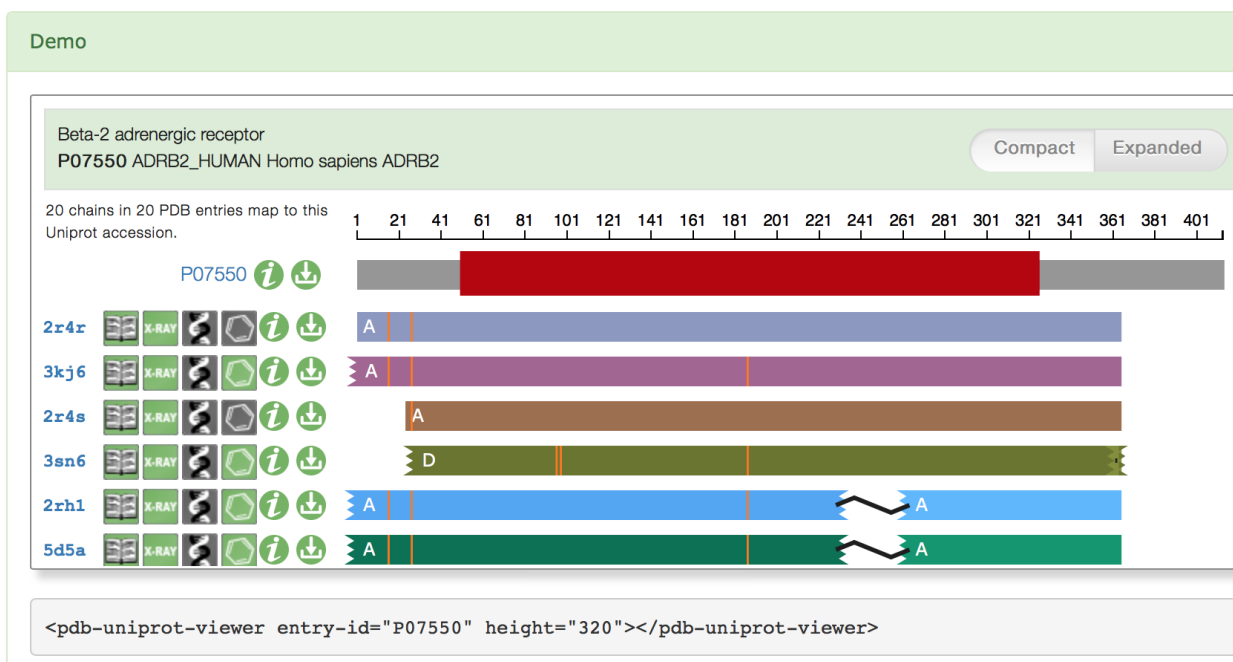

Figure 5: A demo of the PDB UniProt Viewer component

- **PDB_REDO**

The PDB_REDO component shows the change in geometric quality (a combined score for Ramachandran plot, side-chain rotamer, and atomic packing quality) and fit to the experimental data between original PDB entry and its re-refined and rebuilt PDB_REDO counterpart.
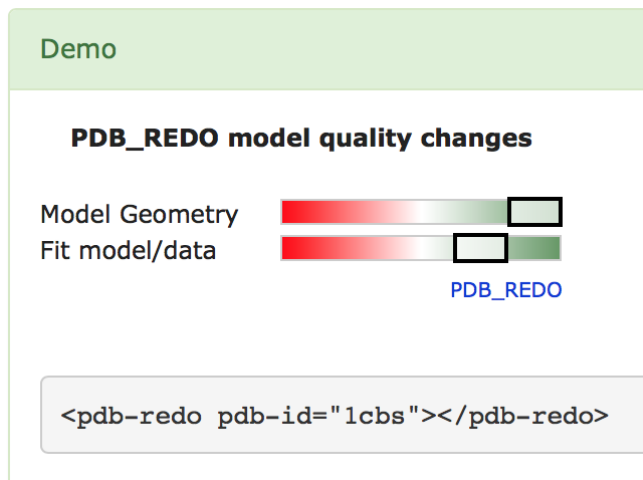

Figure 6: A demo of the PDB_REDO component

- **PDB Residue Interactions**

This is a component based on the residue contacts viewer in Rajini. It displays, in graphical form, the atomic contacts between each of the secondary structural elements (helices and sheets) in a protein. The broader the connection between each of the secondary structural elements, the more atomic contacts are involved in the interface. Hovering over these connections will display the exact number of atomic contacts in this interaction.
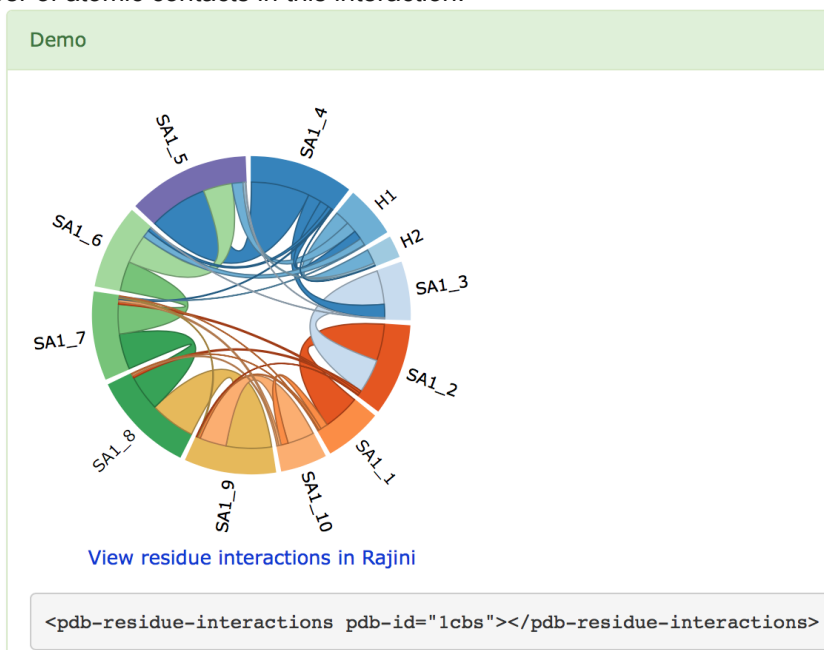

Figure 7: A demo of the PDB Residue Interactions component

- **PDB 3D Complex**

The PDB 3D complex is a server that analyses the complexes in crystal structures. This component gives a compact summary of the results for a particular PDB code and assembly. Selecting 'more details' gives further information based on the 3D Complex prediction and clicking on the 'Evidence' text links out to the server for further information.

Figure 8: A demo of the PDB 3D Complex component

- **LiteMol**

LiteMol is a streamlined structure viewer which enables a PDB structure to be explored within a browser rather than requiring pre-installed molecular graphics software. Navigation is simple, with rotation of the camera using the left mouse button, zooming controlled with the right mouse button and clicking on a residue or atom to center the view to this point. For structures determined by X-ray crystallography, there is also the option to view electron density of the structure where structure factors have been deposited to the PDB. Litemol also displays validation and domain information for the structure.
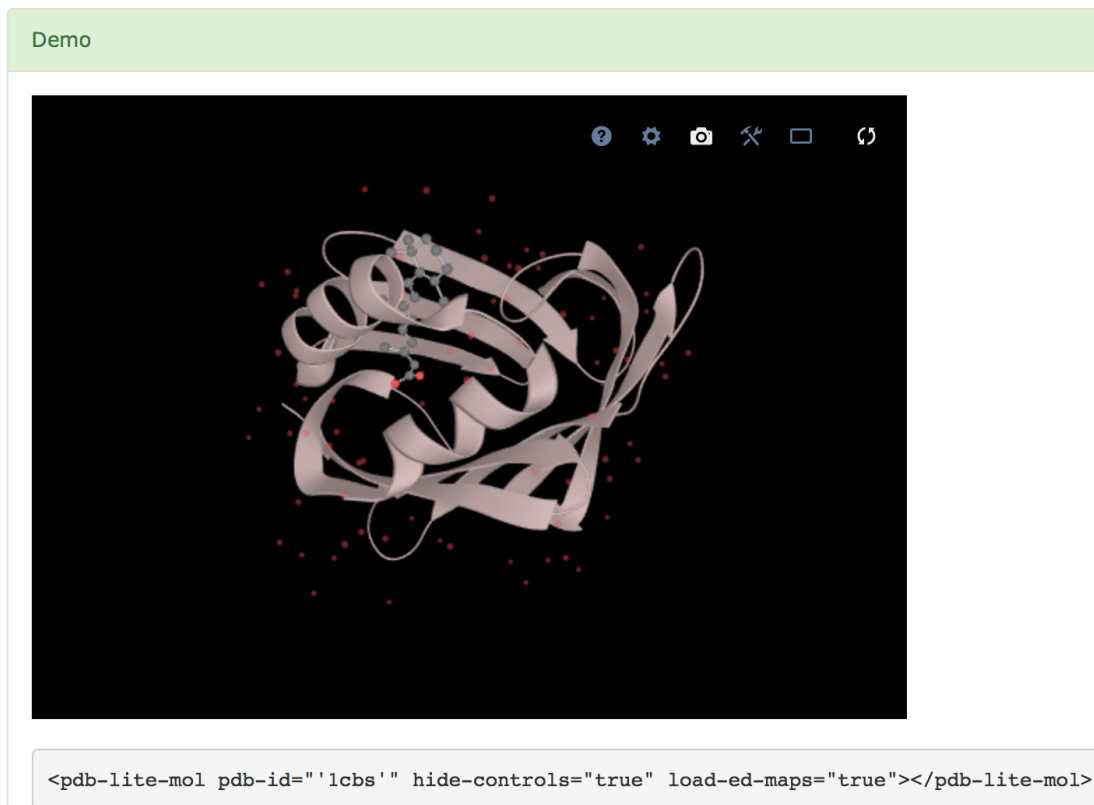
Figure 9: A demo of the LiteMol component

In the Virtual Folder demonstration, the user can specify either a PDB ID or a UniProt ID they are interested in. Following this, only components relevant to that specific ID are shown as a list of collapsible items (See Figure 10). Users can specify a second ID that will be appended to the list of ID's (See Figure 11). Users can choose to delete the ID's they are no longer interested in.
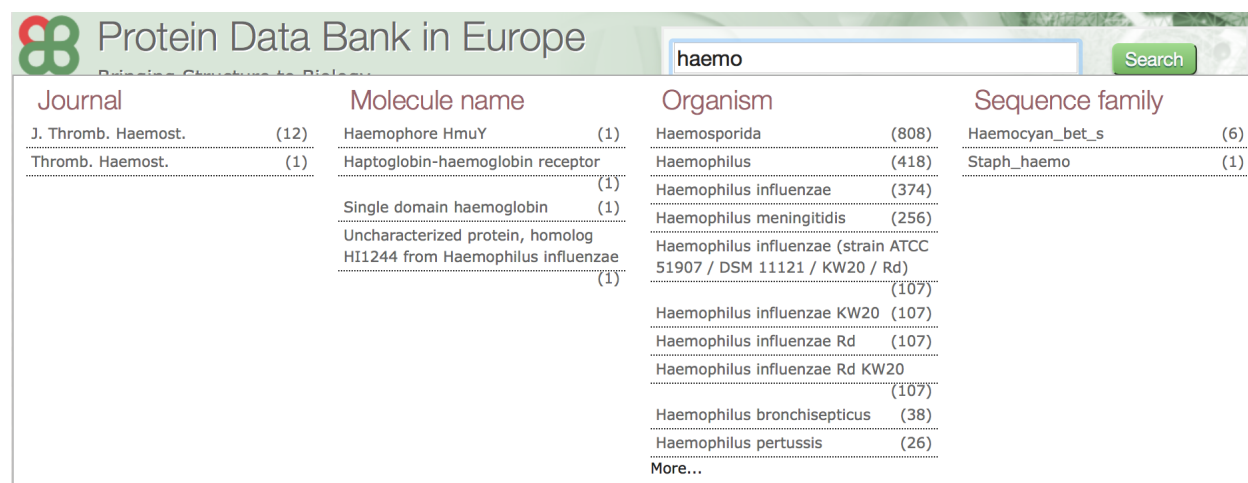
Figure 10: The dataset demo page with 1cbs as an example. Only PDB Prints and PDB_REDO components are shown here, the rest of the components are collapsed.

Figure 11: The dataset demo page with 1cbs and 1atp as examples. When 1atp is specified, the collapsible components for 1atp (here all the components are collapsed) are appended on top of 1cbs.

3.2 Integration of PDBE search API

The Virtual Folder currently contains a page that allows users to use the PDBe autocomplete feature with basic functionality. Tomas Kulhanek (STFC) has adapted the PDB autocomplete feature to use the framework used by the Virtual Folder web ui (Aurelia). The feature utilizes the PDBe Search API http://www.ebi.ac.uk/pdbe/search/pdb-autocomplete/select.

The autocomplete search bar pops out a suggestions when a user starts typing in it. The suggestions displays the different terms that match the input and the number of PDB entries containing each term.

Figure 11: The autocomplete feature with 'haemo' typed as an example, which pops out suggestions for the user.

Delivery and Schedule

This deliverable was delivered on schedule.

Background information

Objectives

This WP is centered on building and operating the VRE web portal that will provide the entry point for users, developers and all other stakeholders. We will build a web portal integrating all already existing and operating services from the various partners and the WeNMR Virtual Research Community (O5.1), and expand it to include new portals, training material and knowledge, and a support center (O5.2, O5.3). In order to better serve the community, customized end-user VMs and/or application containers (e.g. via Docker) will be built for various scenarios (O5.4), to be used on local infrastructures (e.g. within a company) or on the EGI federated cloud resources. Additionally, portals for newly identified applications will be developed and put in production during the project to increase the service portfolio of the VRE (O5.5). The list of objectives is thus:

• O5.1: Deployment and operation of the West-Life-VRE portal, integrating all relevant existing services, training and support components (from WeNMR and other partner sites) and extending them.

• O5.2: Establishment and operation of the West-Life-VRE support and expertise center for users and software developers, covering all VRE areas. This task will cooperate closely with the relevant EGI-Engage Competence Centers (e.g. MoBrain).

• O5.3: Provision of information and training material covering all VRE areas and offered

services.

• O5.4: Development and integration of new service portals.

• O5.5: Provision of customized end-users VMs and/or containers for various applications.

Description of work and role of participants

The above objectives will be addressed through the following tasks:

Task 5.1 – Deployment and operation of the West-Life portal (Luna, all). This task will directly address O5.1. It will start by defining the baseline of existing services across all partners (such as X-ray crystallography from CCP4 and the corresponding ones for cryoEM from the CSIC) together with those of the WeNMR VRC. The CSIC will contribute with the Web Services developed at the Instruct Image Processing Center in Madrid, making use of the Web interface of the SCIPION platform for software integration. These will then be integrated into a new VRE portal which will provide end users with a friendly and dynamical entry point to all services, knowledge and support center. The portal will be built on innovative technology developed by LUNA and we aim to West-Life migrate when possible existing portals to make direct use of the

technology solutions offered by LUNA. In this task, we will also investigate and harmonize user authentication and authorization mechanisms (AAI) (e.g. both the Instruct and the WeNMR sites have user registration mechanisms in place, and WeNMR has implemented a single-sign-on (SSO) mechanism connected to Edugain). The choice and implementation of AAI mechanism will be done in close collaboration with EGI-Engage to maximize compatibility and impact. The new VRE portal will also implement tools and services related to data discovery and access (see WP6).

Task 5.2 – Knowledge and support center (Instruct, all). This task will directly address O5.2 and O5.3. We will integrate the existing knowledge and support center of WeNMR, covering NMR and SAXS services into the new VRE portal, and add all the missing components (tutorials, use cases, help center) to support X-ray crystallography, cryo-electron microscopy and the related integrative methods. A choice will have to be made early on in the project for technology platform to build this knowledge and support center, since various existing components currently use different solutions (e.g. the Instruct web site is based on php while WeNMR operates on Drupal). As in Task 5.1, this will be done in close collaboration with the related EGI-Engage Competence Centers to minimize heterogeneity and maximize impact. Again, in this task, we will as much as possible built on the integrated solutions developed by LUNA.

Task 5.3 – Development and integration of new service portals (UU, all). This task will directly address O5.5. While most of the existing WeNMR portals are already making use of the EGI Grid infrastructure with support from several NGIs within and outside Europe, this VRE project will be adding several portals that are already in place but depend on local and possibly limited resources, as is currently the case for most services for X-ray crystallography and cryoEM. This task will interface those portals (and newly identified ones during the projects) to the most suited e-Infrastructure solution(s), being it grid, CLOUD or HPC resources. Note that we will benefit here from the interaction with various Competence Centers under the new EGI-Engage project, specially the MoBrain Competence Center, to which several partners of West-Life VRE participate (UU,CSIC, CIRMMP and STFC) . Care will be taken to offer user-friendly interfaces, with a VRE-integrated AAI. The most suited submission mechanisms will be selected. For example, we might adopt the efficient DIRAC4EGI service, but could also build on CLOUD and desktop grid (crowd computing) resources offered by the International Desktop Grid Federation (IDGF). A commercial service will also be offered by LUNA for users (both for profit and non-profit) requesting priority access to resources.

Task 5.4 – Customized end-users VMs (STFC, all). This task will directly address O5.5. Structural biology research has been targeting increasingly larger macromolecular machinery of the cell. Consequently, researchers need access to a wide range of techniques and expertise in order to truly exploit structural biology data. In most cases, however they are expert in only one or a few techniques and

associated software. In this task we will build custom VMs for different use cases, with all the necessary software, documentation and examples. Thanks to their suitably designed customization, these VMs will be useful not only to expert structural biologists but also to researchers who want to exploit structural biology as a tool to gain insight in their biological/biomedical research. Different VM types and/or application containers (e.g. via Docker) will be provided, to allow use on both the EGI Federated Cloud and OpenStack/Nebula resources for example, but also local installation on a user's laptop (e.g. with VirtualBox and VMware). This will also potentially be an attractive mechanism for offering commercial services to companies, on their own internal infrastructure when IP issues are preventing external use.

Deliverables

Number	Name	Lead	Type	Dissemination	Delivery date
D5.1	Project portal	Luna	DEC	PU	3
D5.2	Overview (baseline) of services and portals to be integrated into the new VRE	UU	R	PU	4
D5.3	Prototype of the new VRE portal functionality	LUNA	DEC	CO	6
D5.4	Report on activities of the Helpdesk	INSTRUCT	R	PU	18
D5.5	VRE-integrated PDBe search and query API's	EMBL	R	PU	18
D5.6	Report on available VMs with associated documentation/use case for each of them	STFC	R	PU	24
D5.7	Report on access and usage statistics of the various services	UU	R	PU	24
D5.8	Report on access and usage statistics of the various services	UU	R	PU	36
D5.9	Update Report on activities of the Helpdesk	INSTRUCT	R	PU	36

West-Life