`

Project Number: 675858
Project Acronym: West-Life
Project title: World-wide E-infrastructure for structural biology

Periodic Technical Report
Part B

Period covered by the report: from 1 Nov 2015 to 31 March 2017
Periodic report: 1st

Contents

1. Explanation of the work carried out by the beneficiaries and Overview of the progress

1.1 Objectives

The overarching objective of this project is to bring the world of complex data analysis in Structural Biology to a simple Web browser-based Virtual Research Environment (VRE), available to any laboratory involved in the experimental structural characterization of biomolecules and their complexes and assemblies. This general objective is being addressed in detail along the several concrete sub objectives described below.

Objective 1: Provide analysis solutions for the different Structural Biology approaches

An initial version of the West-Life portal was delivered in M3 (D5.1) and has been continuously improved since then. A report on the previous level of activity of partners' services supported us in making plans and tracking progress as reported below (D5.2).

A new implementation of the PDB-REDO service is four times faster (D7.1). This facilitates the development of integrated pipelines that include PDB-REDO enhancement of the structure as one step. The ProteinCCD construct design service has also been enhanced (D7.2). There have been enhancements to WeNMR services and to the Scipion EM service as described below, and new services have been put into operation (e.g. the DISVIS and POWERFIT web portals).

Objective 2: Provide automated pipelines to handle multi-technique datasets in an integrative manner

A common design for security (D4.2) has facilitated progress towards real Single Sign On for European structural biologists. Similarly, a design document for job submission mechanisms that interact with the evolving e-Infrastructure has supported progress in partner services, with compatible mechanisms that allow pipelining. The WP4 report below details the progress made (D4.3).

D5.5 provided web components to integrate search of the PDBe. Since many structural services reuse PDB entries, as these new components are taken up this will enhance many web sites used by structural biologists. Importantly, these components integrate access to PDB-REDO, so making it easier to use the enhanced structures rather than the original deposition.

Objective 3: Provide integrated data management for single and multi-technique projects, based on existing e-infrastructure

The interactions between West-Life partners are resulting in the development of combined pipelines for structural analysis. It is now possible to submit files developed with the Scipion Web Tools to POWERFIT for refinement. Similarly, it is possible to submit from FANTEN to HADDOCK. We have implemented the transfer of crystallographic data files between the CCP4Online, ARP/wARP and PDB-REDO services, and are working on automated protocols that take into account data quality.

We have built a virtual folder view of scattered data. It provides a consistent view of the files for a research project, regardless of the experimental facility in which they were obtained and regardless of their current location. This view is available through a web interface, and also as a mounted file system for access by programs which the user runs (D6.1). The further plans for this are set out in D4.4.

This work uses the B2DROP service provided by EUDAT, and also integrates other repositories like DropBox. In order to facilitate use of EUDAT services by structural biologists, we worked with ARIA and EUDAT to ensure that Instruct userids are accepted by EUDAT's authentication service B2ACCESS [https://www.structuralbiology.eu/update/news/publish-data-with-b2share/]. Instruct, a partner in West-Life, is the ESFRI infrastructure for structural biology. Its member countries are BE, CZ, DK, ES, FR, IL, IT, NL, PT, SK, and UK. Soon, these credentials will also be accepted by B2DROP – at the time of writing, they are already accepted by B2SHARE.

The output from this work is available as a Virtual Machine [VM] suitable for running locally and also for use on EGI resources.

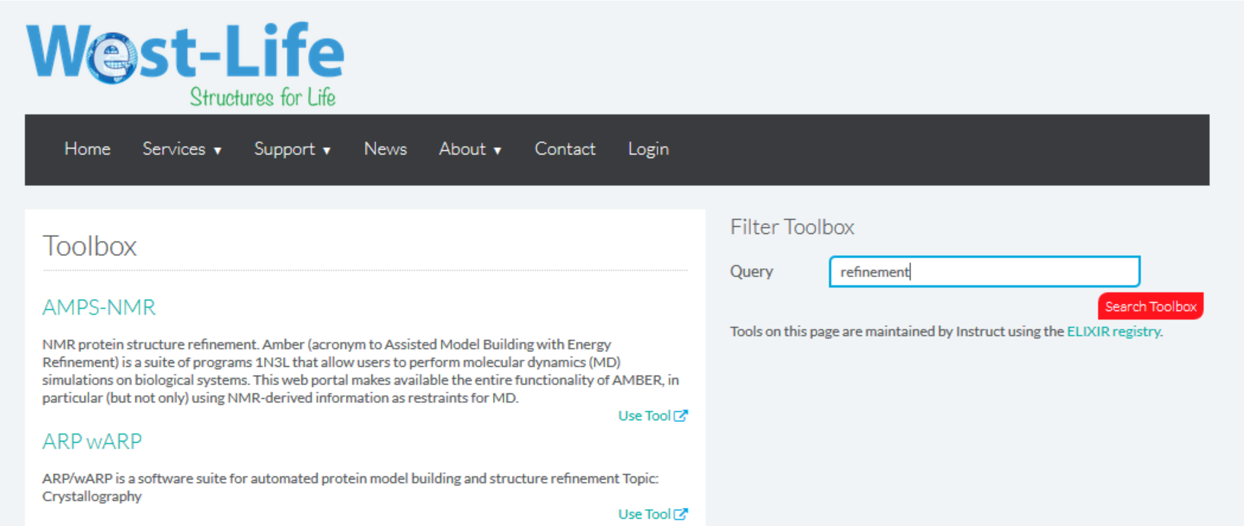
Objective 4: Foster best practices, collaboration and training of end users

We have provided and/or participated in five training events for structural biologists, our end users (D2.1). Further, the West-Life portal contains and aggregates links to, training material for structural biologists (D1.2).

We organised or participated in four training events for developers (D2.2).

West-Life organised round table meeting, jointly with iNext, involving representatives from large ESFRI and infrastructure projects like IMI to understand the requirements from these communities (D3.2).

This activity has informed our sustainability plan (D1.4).



Toolbox

AMPS-NMR

NMR protein structure refinement. Amber (acronym to Assisted Model Building with Energy Refinement) is a suite of programs 1N3L that allow users to perform molecular dynamics (MD) simulations on biological systems. This web portal makes available the entire functionality of AMBER, in particular (but not only) using NMR-derived information as restraints for MD.

ARP wARP

ARP/wARP is a software suite for automated protein model building and structure refinement Topic: Crystallography

Tools on this page are maintained by Instruct using the ELIXIR registry.

Fig 1.1 Part of the West-Life web site

1.2 Explanation of the work carried per WP

1.2.1 Work Package 1 Management

1.2.1.1 Summary and significant results

This work package is concerned with management of the West-Life project, communication and reporting. The project is on course for successful completion, and we have begun the development of a sustainability plan to maintain the VRE beyond the end of the current project.

1.2.1.2 Objectives and Tasks

The objectives of the management work package are closely aligned to the following tasks:

Task 1.1: To project manage West-Life during Years 1-3

Task 1.2: To manage project meetings including the Annual General Meetings for all partners, the Executive Steering Committee meetings and all other governance and operational meetings for the project.

Task 1.3: To manage communication channels between partners, the Coordinator and the EC project officer, including the provision and operation of a project website. To manage collaborative relationships with industrial partners and third parties.

Task 1.4: To manage all reporting to the Commission including financial and WP reports.

Task 1.5: To develop a sustainability plan for the VRE that will see its broad adoption to service a user community spanning from the public to specialist.

1.2.1.3 Work carried out in P1

Task 1.1: Overall coordination of the project

Coordination of the work is achieved through the committees as described below. Also there is a mailing list provided by partner MU, and a wiki provided by partner LUNA.

Developers from all partners met at a Hackathon on 19-20 July 2016 in Paris, to add content to the project web site.

There was an All Partner Meeting in Abingdon, UK to begin the collaboration in January 2016, and a second one in Florence in January 2017. The latter was followed by a workshop for developers, led by Andrei Tsaregorodtsev of EGI.

In addition the work package leaders arrange teleconferences and visits as required.

Task 1.2: Formation and management of Project Boards and Committees

The kick off meeting in January 2016 appointed an Executive Steering Committee chaired by Alexandre Bonvin (Utrecht University) and consisting of Martyn Winn (STFC), Ales Krenek (MU), Jose Maria Carazo (CSIC), and Lucia Banci (CERM). This met on 27 April 2016, and held a series of online meetings in the first quarter of 2017 to discuss recommendations from the Scientific Advisory Board.

All work package leaders participate in a monthly Operations Board telephone conference to coordinate the work. A small Project Management Board also meets monthly between meetings of the Ops Board.

The Scientific Advisory Board consists of Arwen Pearson (CFEL), Erwin Laure (KTH, BioExcel), and Piotr Sliz (Harvard, SBGrid). They attended the second All Partner Meeting in January 2017.

Task 1.3: Communication

Project members have participated in conferences and training workshops of the structural biology community, and of e-infrastructure providers, as reported in the WP2 report below. Outreach is also supported by the project web site http://west-life.eu.

In order to communicate with potential industrial partners, we attended the conference Protein Structure Determination in Industry in 2016, and will attend it again in 2017.

Task 1.4: Reporting to the Commission

PI Martyn Winn, Project Manager Chris Morris, and Antonio Rosato met with the Project Officer Victoria Tsoukala on 5th October 2016 to provide an informal update on the West-Life project. Other reporting has been via the web site SyGMA. An amendment to the grant agreement was submitted as described in 1.2.1.5 below.

Task 1.5: Development of a sustainability plan (STFC, all)

D1.4 presented initial steps towards a sustainability plan, and planned the necessary actions required to complete it. In addition, some services of WestLife (namely the WeNMR ones that have a long history of working with European e-infrastructures) have been included as thematic services in the EOSC-Hub proposal that was submitted by EGI/EUDAT/INDIGO-Datacloud. In addition, partner CSIC has presented a proposed to the EOSCpilot scheme for work complementary to the current grant on metadata for Electron Microscopy workflows, which will open the way to integration of EM processing into the European Open Science Cloud.

1.2.1.4 Progress towards Deliverables and Milestones

Number	**Title**	**Lead**	**Type /planned delivery date**	**Progress**
D1.1	Kick off meeting	STFC	Public report Month 3	Delivered
D1.2	Project web pages	Instruct	Other public Month 3	Delivered See https://west-life.eu
D1.3	Periodic report	STFC	Public report	Delivered This report
D1.4	Draft sustainability work plan	Instruct		Delivered
D1.5	Sustainability report	Instruct	Public report Month 36	Reporting Period 2
D1.6	Final project conference	STFC	Public report Month 36	Reporting Period 2 Reconsidered, see 1.2.1.5 below
D1.7	Final Report	STFC	Public Report Month 36	Reporting Period 2

Milestone Number	Milestone Name	Estimated Date	Means of Verification
MS1	Kick off meeting	3 / Delivered	Report disseminated
MS2	Web site launched	3 / Delivered	Website up and running with content in place
MS3	First periodic report	18/ Delivered	Report submitted
MS4	Sustainability report approved by Executive Steering Committee	M30	Reporting Period 2

1.2.1.5 Deviations from Annex 1

The dates of a few deliverables and milestones required adjustment for reasons reported elsewhere in this document.

We now consider that it will be more effective to hold the West-Life conference in 2018Q1, rather than in the final month of the project, so it can plan achievements rather than merely celebrate them.

Partner Instruct does not employ staff directly, but instead is operated by people formally employed by the University of Oxford. The University must therefore be recognised as a third party in the activity.

Partner CIRMMP had the opportunity to obtain a contribution to the project from someone employed by the University of Florence, which should therefore also be recognised as a third party.

These issues were all reported in an amendment to the GA, which was approved on 19th May 2017.

1.2.2 Work package 2 Dissemination, Training, and Outreach

1.2.2.1 Summary and significant results

This work package aims to disseminate the achievements of West-Life within and outside the existing structural biology community; with a special focus on outreach to training and engaging with new user communities and industry. Over the reporting period of 18 months, the work package has organised 5 workshops and 5 training events collaborating with other projects and networks as appropriate. A social media presence has been established and West-Life has been represented in a wide variety of platforms and will continue to do the same to publicise the contribution of the West-Life VRE to the scientific communities around the world.

1.2.2.2 Objectives and Tasks

Objective 2.1: Training new and existing communities of users in the tools, services and methods offered by the West-Life VRE to achieve high impact research outcomes
Objective 2.2: Enabling developers to provide new solutions for currently intractable problems that limit integrative approaches in structural biology
Objective 2.3: Engaging all sectors from the public to research specialists and communicating the power of a virtual research environment to provide a seamless resource for the management, analysis, interpretation, archiving and protection of data acquired from structural biology methods

Task 2.1 – Training programmes: user communities (Instruct, all).

This task addresses Objective 2.1 and implements a series of training events. Two training events in each of years 1, 2 and 3 will be tailored to the existing user structural biology community and provide information and practical experience of data handling, including the availability, access and use of existing tools and the data outcomes. The user workshops interface with existing Instruct Training courses, which provide practical training on new and integrative structural biology technologies and include best practices in data acquisition and handling at the research 'bench'. One training event per year (in years 1 - 3) will deal with archiving data, existing common data services (EUDAT http://www.eudat.eu), charters (Access charter). A programme for training components of West-Life has been developed by month 9, incorporating shared training events with stakeholders where relevant (M2.1).

Training events may include a virtual web-based component and face-to-face meetings. These will combine user groups and include a 'user clinic' at which practical management of problems, outstanding requirements and possible solutions can be discussed.

Task 2.2 – Training programmes: developers (Instruct, all).

This task addresses Objective 2.2 by implementing a programme of training events for developers, which builds partly on the feedback from the user communities on problems and gaps in integrative resources and partly on the known issues in developing a full suite of integrated services. More specifically, these training events aimed at developers will focus on the solutions and on the presentation of the opportunities that West-Life has deployed and made available to the entire community of software developers and providers of computational services. By participating in the present training events, developers both within and outside the present partnership will have access to expertise, information, and guidelines to make their tools available via the West-Life platform and/or to exploit the computational resources gathered within the present project. This may lead to an expansion of services available to our users beyond the tools created by the partners.

The programme will implement at least three training workshops (one per year).

Task 2.3 – Dissemination and outreach (Instruct, all).

A key part of delivering services to the community at large is to promote understanding and importance of structural biology by providing information on the value of the research being supported, the resources underpinning the service and to share knowledge and experiences arising from the project with a broad stakeholder community. In specific terms, this task addresses Objective 2.3 and is devoted to developing and strengthening interactions with existing and new communities bridging the academic, industry, policy and decision makers and public sectors.

Implementation of this task will address the three target groups separately and also bring all parties together for the Kick-off meeting and the end of project meeting.

1.2.2.3 Work carried out in P1

Task 2.1 – Training programmes for user communities

In the initial reporting period of 18 months, West-Life organised 5 different training events targeting user communities in Europe and worldwide, when the workshop was organised in collaboration, the partners are indicated. A summary of the training events is presented as a table below:

Title	Organiser	Country	Audience	Number of participants	Countries represented
Second Advanced methods for the integration of diverse structural data with NMR data	UU (Partner 8)	Netherlands	Structural Biology community	20	Czech Republic, France, Germany, Italy, Netherlands, Singapore, Spain, UK, USA
First I2PC-FEI "hands on" course on image processing applied to the structural characterization of biological macromolecules	CNB-CSIC (Partner 5), FEI	Spain	Structural Biology community	20	Austria, Czech Republic, Denmark, France, Germany, Netherlands, Italy, Spain, UK
Cryo EM Solving the Structure of Macromolecular Complexes: A Hands-on Workshop at International Symposium on Grids and Clouds 2016	Academia Sinica Grid Computing Centre (ASGC), UU (Partner 8), CNB-CSIC (Partner 5)	Taipei	New user community	20	India, Malaysia, Netherlands, Spain, Taiwan
From 2D images to 3D structures: A practical course on Electron Microscopy Single Particle Analysis	CNB-CSIC (Partner 5), STFC (Partner 1), EMBL-EBI (Partner 3)	Spain	Structural Biology community	22	Czech Republic, Slovenia, Denmark, UK, France, Netherlands, Germany, Switzerland, Latvia, Italy, Spain
Cryo-EM Workshop at International Symposium on Grids and Clouds 2017	Academia Sinica Grid Computing Centre (ASGC), UU (Partner 8), CNB-CSIC (Partner 5)	Taipei	New user community	40	Czech Republic, France, India, Italy, Malaysia, Netherlands, Spain, Taiwan

The workshop participants represented both members from existing structural biology communities and new user communities. The events were focused on providing practical knowledge of using scientific techniques and also, using tools and services made available through West-Life, including image processing and data analysis of structural data.

West-Life training contributions are not just restricted to events organised by the project. Scientists from the project partners conduct sessions at workshops or courses organised by collaborators. These sessions are aimed at training users in specific methodologies/tools/services enabled through West-Life. Thirteen such exclusively workshop/tutorial sessions (we do not count here presentations at general scientific conferences) were provided by West-Life partners. Conducting training sessions at such events help the dissemination of West-Life project and tools/services enabled through the project to a worldwide community of users.

Task 2.2 – Training programmes for developers

There were 4 developer training events organised during the initial 18 months of the project. The objectives of the training workshops included sharing expertise within and outside the West-Life partners, including building bridges with tools/services provided by e-Infrastructures such as EUDAT (https://eudat.eu/services) and EGI (https://www.egi.eu). Courses also included approaches of cloud infrastructure setup and using software for distributed computing. A summary of the courses enabled through West-Life is below:

Title	**Organiser**	**Country**	**Audience**	**Number of participants**	**Countries represented**
West-Life Hackathon	Luna (Partner 9), STFC (Partner 1)	France	West-Life Developers	10	Czech Republic, France, Italy, Netherlands, Spain, UK
DIRAC Workshop for West-Life	CIRMMP (Partner 6)	Italy	West-Life Developers	12	Czech Republic, Italy, Netherlands, Spain, UK
Hands-On course on Cloud Orchestration	MU (Partner 4)	Czech Republic	Developer community	9	Czech Republic, Germany, Greece, Netherlands, Romania, Spain, UK
Developer conference at Digital Infrastructure for Research September 2016	EGI, EUDAT, RDA, STFC (Partner 1)	Poland	e-Infrastructure Developer community	271	European countries

West-Life also was represented at developer conferences such as DI4R along with programmers and services providers from various e-Infrastructures and industrial stakeholders. These events present opportunity to partnership developers to gain information and expertise in scientific developments and also to disseminate information on access and availability of tools and services that West-Life partners provide.

It was agreed that a part of the future developer training workshops will include web-based training events such as webinars.

Task 2.3 – Dissemination and outreach

Keeping the website current with up-to-date information and dynamic content is imperative for engaging with user communities. The West-Life website (https://west-life.eu/) has been set up in collaboration with Instruct to share news and events which increases the reach to thousands of Instruct users (https://www.structuralbiology.eu/) in addition to disseminating important structural biology information to West-Life users. The website also has a Twitter widget engaging with different user communities, partners and collaborators. The @WestLifeSB twitter account has 120 followers and has 520 tweets engaging with policy makers, partners, collaborators and users, averaging excess of 100 profile views per month. Owing to the emergence of Twitter as a support platform, with specific instance of such support request in West-Life, the website also encourages users to contact West-Life using the twitter handle. Webinars and lecture, collected by WeNMR project, on the tools and services available to the users through West-Life are uploaded to the WeNMR YouTube Channel. More lectures from workshops will be added to this channel for further dissemination.

The West-Life website has also been improved with the facility of a forum facility for immediate support of user communities. Where the tools and services provided through West-Life have their own established portals and support mechanisms, the forums and associated support pages in West-Life website will direct users to the correct help centre.

Work Package 2 produced a West-Life flyer after consultation with partners and it is made available for everyone in the project to circulate among collaborators and at events. The flyer contents and the design were revisited through consultation with new user communities including a public engagement expert, to make sure that the message and the text remain relevant not just to the structural biology community but to a wider audience. The work package also produced template materials for presentations and posters for use by partners when representing the project at different events.

As an e-Infrastructure project, West-Life will also benefit from continued presence at conferences and workshops for better dissemination of information about the project and the VRE. Partners of West-Life presented 5 posters at different conferences and cited West-Life at 60 lectures at conferences worldwide. West-Life was also acknowledged in 10 published journal articles. West-Life was also represented through presence at different events with a list of future events documented as part of the engagement plan available in the internal wiki: http://internal-wiki.west-life.eu/w/index.php?title=Engagement_Plan. Project partners are continually encouraged to use West-Life branded templates for presentations and posters, use West-Life logos, namecheck other partners and distribute business cards when attending networking events.

West-Life also co-organised CryoEM Best Practices workshop along with Instruct. The workshop was targeted to Industry and academic representatives sharing ideas and practices to address some of the concerns of the cryoEM part of the structural biology community.

Title	**Organiser**	**Country**	**Audience**	**Number of participants**	**Countries represented**
Instruct workshop on CryoEM Best Practices	Instruct (Partner 7), CSIC (Partner 5)	United Kingdom	Industry, Structural Biology community	30	Czech Republic, France, Germany, Netherlands, Spain, UK, USA

Engaging with new user communities such as Biomedical Sciences Research Infrastructures is critical, particularly when structural biology data is becoming relevant and necessary for translational research. As part of this commitment, the foresight meeting from milestone MS7 was co-organised in conjunction with the Round Table meeting being organised by WP3 as part of deliverable D3.2. The Meeting held along with the WP3 Roundtable meeting in advance of the Instruct Biennial Structural Biology Meeting on 24th May at Brno, Czech Republic was the culmination of networking activities lead by Instruct in collaboration with other Biomedical Research Infrastructures. In addition to all the consortium meetings taken place in the frame of CORBEL the following meetings provided background and networking opportunities for the BMS Research Infrastructure community:

- ARBRE-MOBIEU 2017 plenary meeting, Porto, Portugal - 21th/25th March 2017 - Claudia Alen Amaro
- 2nd ESFRI Exchange of Experience Workshop, Malaga, Spain - 18th January 2017 - Claudia Alen Amaro
- RI Long term sustainability stakeholders workshop - 25th November 2016 - Lucia Banci, Ondrej Hradil
- CORBEL Medical Users Infrastructure Forum, Paris, France - 12th October 2016 - Susan Daenke
- CORBEL Annual General Meeting, Graz, Austria - 18th/19th October 2016 - Antonio Rosato, Narayanan Krishnan
- International Conference on Research Infrastructure (ICRI) 2016, Cape Town, South Africa - 3rd/5th October 2016 - Susan Daenke

1.2.2.4 Progress towards Deliverables and Milestones

Number	**Title**	**Lead**	**Type /planned delivery date**	**Progress**
D2.1	Interim report on T2.1	7 - Instruct	Public report / 30th April 2017	Delivered
D2.2	Interim report on T2.2	7 - Instruct	Public report / 30th April 2017	Delivered
D2.3	Summary report of SB community engagement	7 - Instruct	Public report / 30th April 2017	Delivered
D2.4	Report on industrial user engagement	7 - Instruct	Public report / 30th October 2017	24 - Workshops and other engagement events being held along with Industrial partners will form part of the deliverable
D2.5	Engagement report	7 – Instruct	Public report / 30th October 2018	36 - All the engagement initiatives taken through the life of the project will be summarised in the report with a view of sustainability of the achievements

Milestone Number	Milestone Name	Estimated Date	Means of Verification
MS5	Training programme plan	9 / Delivered	Schedule for training programmes years 1-3 approved and published on the website
MS6	Engagement plan	12 / Delivered	Outline plan and schedule for Engagement approved and published on the website
MS7	Foresight Meeting	15 / Delivered	Meeting held along with the WP3 Roundtable meeting at Instruct Biennial Structural Biology Meeting on 24th May at Brno, Czech Republic.

1.2.2.5 Deviations from Annex 1

None

1.2.3 Work package 3: Networking

The overall goal of this WP is to maximise communication and identify opportunities for collaboration with European and global infrastructures that may leverage Structural Biology data as well as with European and global initiatives addressing scientific data, such as the Research Data Alliance and other e-Infrastructures. A further objective of this WP is to interact with organizations involved in the development of science policy, including national and European scientific academies, funding bodies and associations thereof.

In practice, we aim to evaluate the needs expressed by regional, national, European or global infrastructures not directly involved in the present partnership, with regard to the availability and use of structural biology data. In this respect, Instruct has a central role by leveraging its close relationship with the other BMS RIs to foster the interaction between the large ESFRI community and West-Life. The Interest Group in Structural Biology within the Research Data Alliance (RDA), which is chaired by Lucia Banci (Partner CIRMMP), Chris Morris (Partner STFC) and Antonio Rosato (Partner CIRMMP), provides a route for interaction with data initiatives. The RDA is a global forum that comprehensively addresses all aspects of data science.

1.2.3.1 Summary and significant results

This WP has produced a deliverable, D3.1, entitled “Assessment of the life cycle of structural data and comparison with other scientific data”. This report was based on the data life cycle as defined by the UK Data Archive, which identifies six stages of the cycle (creating data, processing data, analysing data, preserving data, giving access to data, re-using data). The report addressed structural biology data in the context of these steps and added a final stage to the life cycle: “discarding data”. Further activities of WP3 involved the organization of a Round Table, entitled "Bringing together the biomedical scientific communities: the role of research infrastructures". The Round Table brought together representatives of many different biomedical ESFRIs, including ELIXIR, EATRIS, ECRIN, EuOpenScreen, EuroBioImaging and ISBE.

This WP is also surveying the scientific community regarding their usage of structural biology data and their perception of existing gaps at the level of data standards and software availability for a broader usage of such data. The results of this survey could contribute to the discussion at the aforementioned round table.

Finally, we organized a session of the Interest Group in Structural Biology (SBIG) focused on the life cycle of structural biology data on the occasion of the Ninth RDA Plenary (https://www.rd-alliance.org/ig-structural-biology-rda-9th-plenary-meeting). This session resolved to submit its report as a candidate RDA Supporting Output.

1.2.3.2 Objectives and Tasks

The detailed objectives of this WP, as specified in the original application, are:

Objective 3.1: identify opportunities for collaboration with European and global infrastructures that may leverage Structural Biology data to enrich/extend its impact on the society at large.

Objective 3.2: identify opportunities for collaboration with European and global initiatives addressing scientific data and other e-Infrastructures

Objective 3.3: interact with organizations involved in the development of science policy

The above objectives are addressed through the following tasks.

Task 3.1 – Interaction with large infrastructures and European projects (CIRMMP, all).

This task addresses Objective 3.1. The present task surveys and reports on requirements and needs expressed by European or global infrastructures not involved in the present partnership, both physical and electronic, with regard to the availability and use of structural biology data. Instruct is also exploiting its existing close relationship with the other BMS RIs in setting up cross-cutting and joint activities that consolidate the interaction between the large ESFRI community and West-Life in order to further develop services with a structural biology focus. The requirements of the various infrastructures are being collected by a combination of the following: surveys, documented direct contacts with key staff and scientists, organization of Round Tables and joint meetings, and/or the implementation of working groups involving representatives of different infrastructures. In addition, “scouting” actions at selected infrastructures (i.e. visits of up to a few days) may be performed, if deemed appropriate. As examples, contacts with the following ESFRI infrastructures have been be sought: EU-Openscreen, ELIXIR, EuroBioImaging, MIRRI, ISBE, EATRIS, EMBRC. Interactions and links will be established also with large European projects, such as the JPIs (notably the JPND on Neurodegenerative Diseases) or institutions, like EMBO. At the global level, we will leverage the connections already well established by the WeNMR project, namely that to the Open Science Grid (OSG) in USA and that to Academia Sinica (AS) in Taiwan, e.g. by co-locating West-Life overseas events at large events organized by OSG or AS respectively.

Task 3.2 – Interaction with data initiatives (STFC, all). This task addresses Objective 3.2.

This task focuses on the involvement of project representatives in European and global fora related to the various aspects of the life cycle of scientific data. From the very beginning of the project, we have leveraged the existence of an Interest Group in Structural Biology, which is chaired by some of the project partners including the coordinator of this WP, within the Research Data Alliance. Furthermore, we are taking other relevant initiatives, from data exchange and data sharing to possible exploitation of archived data. A specific example is that of the European Data Infrastructure, EUDAT. The activities of this task will be tackled through the participation of West-Life partners at relevant events, as well as, when useful, through the establishment of specific task forces/working groups with the aim of identifying synergies and opportunities for a more thorough exploitation of structural data.

Task 3.3 – Interaction with initiatives developing science policy (CIRMMP, all).

This task directly addresses Objective 3.3. This task maintains a close watch on the development of science policies affecting the field of SB in general. We will also evaluate opportunities to establish collaborations with or to bring specific issues/opportunities to the attention of science policy makers or funding bodies. Various project partners are already involved in similar endeavours with a series of

European Institutions and Consortia which identify needs of the various scientific communities and develop strategies to address new directions and challenges for research and innovation. The current Task constitutes a venue where we can compare the corresponding outcomes and develop synergistic strategies.

1.2.3.3 Work carried out in P1

In the period from November 1st 2015 until April 30th, 2017, Deliverable D3.1 was scheduled and submitted. Other deliverables, mainly focusing on the organization of meetings or satellite events of other initiatives, initially planned for the same period instead were postponed due to the lack of suitable opportunities to co-locate them. The rescheduling was agreed with the project PO, and work towards these is now under way and a round table meeting was held, jointly with iNext, involving different ESFRI representatives at Brno, Czech Republic on 24th May. This section describes the work performed so far within the various Tasks.

Task 3.1 – Interaction with large infrastructures and European projects.

The West-Life consortium assessed that the natural counterpart to reach out to in order to fulfill the activities of this Task is the CORBEL partnership (http://www.corbel-project.eu/home.html). CORBEL is an initiative of eleven biological and medical research infrastructures (BMS RIs), aiming to create a platform for harmonised user access to their technological platforms, sample collections and other services. The eleven RIs in CORBEL are listed here: BBMRI-ERIC, EATRIS-ERIC, ECRIN-ERIC, ELIXIR, EMBRC, EURO-BIOIMAGING, EU-OPENSCREEN, INFRAFRONTIER, Instruct, ISBE and MIRRI. Interaction with CORBEL allows our efforts to reach a larger number of BMS RIs.
Instruct is a partner of CORBEL and has been active in various initiatives including, within CORBEL WP5, identifying common processes and standards which harmonize user access across the BMS RIs including authentication and authorization. Instruct also participates in AARC2, working together with national identity federations, research infrastructures, e-infrastructures and libraries to establish the best practices and policies needed to implement interoperable authentication and authorisation for infrastructures (AAIs). Instruct was one of the first of the BMS RIs to develop a web-based access management system (ARIA, aria.structuralbiology.eu/). The ARIA software has been adopted by a number of RIs and CORBEL adopted ARIA for its first open call for research projects. A key component of ARIA is the authentication system which provides a choice of sign-on routes (Instruct, EduGAIN, UMBRELLA) which are connected using custom software which translates the Shibboleth identity to an ARIA identity. The West-Life IdP/SP Proxy will facilitate integration of software services with high-quality AAI methods and assist with eventual migration to ELIXIR. The existing Structural Biology user community will be supported with their existing ARIA identities in each of the West-Life IdP/SP Proxy and ELIXIR AAI. Connected identities are used in the API endpoints to help identify common users between multiple systems and work is underway to integrate identities fully into EUDAT (European cloud data initiative) to allow seamless auto-provisioning.

As part of our endeavor to collect information on the awareness and (potential) usage of structural biology data within and beyond the current community of structural biologists, we had planned to organize an event with other ESFRIs to collect their input. Initially we sought to organize a Round Table on the occasion of the first annual meeting of CORBEL, which took place in October 2016. This would have fulfilled D3.2 and M3.2, due respectively at month 12 (October 2016) and 15 (January 2017), accordingly with Annex 1. Unfortunately, the agenda of the meeting was too crowded to allow for enough time, as needed for an extensive discussion. We consequently scheduled the event in Brno on May 24th, as a special session of the 2nd iNEXT Annual User Meeting, just preceding the 3rd Biennial Scientific Conference of Instruct (Brno, May 24-26). The Round Table, entitled "Bringing together the bio-medical scientific communities: the role of research infrastructures" (https://www.structuralbiology.eu/update/content/bringing-together-the-bio-medical-scientific-communities-the-role-of-research-infrastructures), focussed on the exploitation of Research Infrastructures in the area of Biomedical Sciences, both physical and electronic, by European researchers working in biological and biomedical sciences. This round table meeting aims at developing ideas to foster usage of Research Infrastructures and therefore to increase their impact and role for innovation. In order to provide a specific focus for the discussion, a few selected discussion points have been sent to the participants beforehand.

Key representatives of the following BMS RIs, H2020 projects and European networks have participated as panellists, while other projects such as FRISBI and P4EU were represented in the audience:

RI / Project	**Representative**
EATRIS	David Morrow
ECRIN	Serena Battaglia
Edison (H2020 project)	Steve Brewer
EuOpenScreen	Bahne Stechmann
EuroBioImaging	Antje Keppler
IMI	Hugh Laverty
iNEXT (H2020 project)	Lucia Banci
Instruct	Susan Daenke
ISBE	Vitor Martins Dos Santos
PhenoMeNal (H2020 project)	Merlijn van Rijswijk
West-Life (H2020 project)	Antonio Rosato

Table 1.2.3.2.1

This Round Table fulfils D3.2 and M3.2, as mentioned above, and provided input to D3.3 (Report on requirements by other RIs), which has also been postponed to the same date.

Task 3.2 – Interaction with data initiatives (STFC, all).

As mentioned in the previous section, in this task we focus on the interaction with European and global fora related to the various aspects of the life cycle of scientific data. The first Deliverable – D3.1, a report on the "Assessment of the life cycle of structural data and comparison with other scientific data", was submitted according to the initial schedule, after extensive discussion with all the partners. The report is publicly available from the project web site. STFC collected comments on an initial version of the report also by the participants in the 5th RDA IG/WG Collaboration Meeting, which took place in

Nottingham, UK on June 2016.

The organization of a session of the Interest Group in Structural Biology at one of the RDA plenaries (M3.1) was postponed to M18 (April 2017) as this was the first RDA plenary organized in Europe after the start of the project, while attending a meeting in the USA did not seem beneficial enough to justify the cost to the project. The plenary session took place on April 5th 2017, at 16.00 (https://www.rd-alliance.org/ig-structural-biology-rda-9th-plenary-meeting) and fostered further discussion on the life cycle of structural data. The speakers were:

- John Helliwell, Emeritus Prof of Chemistry, University of Manchester.
- Geerten Vuister, CCPN, Professor of Structural Biology, University of Leicester
- Pablo Conesa, Biocomputing Unit, National Centre for Biotechnology, CSIC
- Sameer Velankar, Team Leader, Protein Data Bank in Europe, EBI
- Antonio Rosato, CERM

The latter three speakers are partners in West-Life; the first two are not. Besides fulfilling M3.1, the additional input should strengthen this document and allow us to publish as an open access article. The meeting empowered the speakers to edit the report and put it forward for approval as an RDA output.

Task 3.3 – Interaction with initiatives developing science policy (CIRMMP, all).

Interactions with key stakeholders at both national and European level are fundamental in order to follow the development of science policy that could affect the SB research community. Three project participants are directly involved in the CORBEL initiative, which has been already mentioned in the context of Task 3.1; this provides a privileged way to monitor science policies that affect the multiple facets of biological sciences. In addition, the role of Lucia Banci as the Italian representative in the "Health & Food" Strategic Working Group of the Strategy Forum on Research Infrastructures and her involvement in the Coordination Committee of the Italian BMS RIs that are landmarks or projects in the ESFRI roadmap was particularly important for the science policy perspective of West-Life. Furthermore, the project INRoad (http://inroad.eu/) recently involved Lucia Banci as member of one external advisory panel (reflection group). INRoad is a Horizon 2020 CSA project that aims at supporting RI policy development in terms of exchange of best practices for national roadmap drafting and evaluation procedures for RI, in order to promote comparability and synchronization of national procedures and harmonization of evaluation mechanisms.

Another Europe-wide channel to expose the West-Life partnership to current developments in science policy is via the participation in the RDA events, both at national and plenary levels. Besides the interactions mentioned in Section 3.2, West-Life partners have taken part in activities such as the various meetings of the "RDA meets researchers" that have been organized throughout Europe (https://www.rd-alliance.org/events.html).

Instruct, as a member of the Medical Infrastructure User Forum (MIUF; part of CORBEL WP3) has worked to promote collaboration across the ESFRI BMS infrastructures by defining priorities on new infrastructure and service development to meet the needs of the scientific user communities. The objective is to build scientific projects / programmes / strategies and promote "cross-border" adoption and use of resources. A survey identified the breadth and depth of knowledge of RI services amongst broad scientific user communities as a starting point for improving RI awareness. The BMS Strategy Forum, a forerunner to MIUF, developed a European Charter for Access to Research Infrastructures - Principles and Guidelines for Access and Related Services, which is now adopted by ESFRI and the European Commission. Research Infrastructures are periodically invited to provide comments on this Charter and suggest possible improvements. Input into policy is also provided through regular participation in RI Exchange of Experience workshops hosted by the European Commission, and initiatives in the development of data sharing policies for clinical and non-clinical data (CORBEL WP3). Further occasions important for understanding how national policies could affect SB, have been Lucia Banci's involvement in various National evaluation committee as, for example, the Building Blocks of Life assessment committee of the Netherlands Organisation for Scientific Research (NWO) and the Italian committee.

The recent nomination of Alexandre Bonvin as scientific advisory board member of the Molecular and Cellular Structure (MCS) cluster at the EMBL-EBI in Hinxton is providing a direct link and input channel with the Protein Data Bank in Europe (PDBe), the major repository of structural data. The first 2017 meeting of the MCS SAB took place on April 25/26. Alexandre Bonvin is further also member of the advisory board of the compute and data infrastructure programme of SURFSara, the main e-Infrastructure provider in the Netherlands.
In the context of large structural biology projects, the role of Jose-Maria Carazo as scientific advisory board member of the European Synchrotron Radiation Facility (ESRF) helps to assure a close link between data producers and data analyzers (like West-Life). Furthermore, in relation to International initiatives, JMC is part of the International Scientific Advisory Committee of the Australian Center for Advanced Molecular Imaging and one of five members of the Top Users Program of the Center for Protein Science of the Chinese Academy of Sciences (at Shanghai).
Among the various meetings related to the above activities, the following are worth mentioning:

- First meeting of the Coordination Committee of the Italian BMS RIs, 20 July 2016, Milan
- La rete nazionale delle Infrastrutture di Ricerca nella roadmap europea: innovazione e competitività in Biomedicina , 28 September 2016, Rome
- ICRI 2016 | Cape Town – South Africa, 3-5 October 2016
- "FAIR Data Management: best practices and open issues", 14 -15 November 2016, Florence Long term sustainability of Research Infrastructures, Exploring RI's full potential - 25 November 2016, Brussels
- 2nd Exchange of Experience Workshop, organized by StR-ESFRI, 18 January 2017 Málaga
- InRoad: Engagement Roadmapping Workshop - 14 March 2017 – Brussels
- ESFRI Roadmap 2018 – giornata informative, 5 April 2017, Rome
- 9th Research Data Alliance Plenary Meeting, Barcelona, Spain, 5-7 April 2017.

On various occasions, the interactions with the above initiatives carried out by the individual West-Life partners were discussed. D3.4 reports the general perception of the project consortium with respect to current science policy in Europe.

1.2.3.4 Progress towards Deliverables and Milestones

Number	Title	Lead	Type /planned delivery date	Progress
D3.1	Assessment of the life cycle of structural data and comparison with other scientific data	STFC	Public report / 31st July 2016	Delivered
D3.2	Organization of a round table or joint meeting involving ESFRIs	INSTRUCT	Public report / Postponed to May 2017	Delivered on May 24th, 2017.
D3.3	Report on requirements by other RI	CIRMMP	Public report / Postponed to May 2017	First draft to be circulated to round table panelists by June 10th
D3.4	Interim report on science policies	CIRMMP	Confidential report / 30th April 2017	Delivered
D3.5	Publication of a joint document on the usage of structural data in different biomedical RI's	CSIC	Public report / 30th October 2017	In preparation
D3.6	Update of requirements by other RI	CIRMMP	Public report / 30th April 2018	Depends on D3.3

Milestone Number	Milestone Name	Estimated Date	Means of Verification
MS8	RDA	Postponed to April 2017	A session of the Interest Group in SB took place at the Ninth RDA Plenary
MS9	Round Table	Delivered on May 24th, 2017.	https://www.structuralbiology.eu/content/bringing-together-the-bio-medical-scientific-communities-the-role-of-research-infrastructures
MS10	Implementation of a Working Group involving infrastructures beyond the current partnership	21	Publication of WG Charter

1.2.3.5 Deviations from Annex 1

Deviations from Annex 1 are described in the Tables of the previous section.

1.2.4 Work Package: 4 Operation

1.2.4.1 Summary and significant results

The main goals of WP4 are keeping the existing services operational, and their gradual transition to the new consolidated architectures. Both the goals are being accomplished according to the project plan. Operation of the existing services proceeds with a sustained number of job submissions to EGI HTC resources (>8 million jobs per year), showing even a moderate increase. The SLA agreement with resource providers has been renewed for 2017 providing 60 million CPU hours on the EGI grid, 60 dedicated core on the EGI Federated Cloud and more than 250 TB storage capacity to support the enmr.eu VO. A consolidated architecture for job submission and portal deployment in general was defined.

1.2.4.2 Objectives and Tasks

The principal objectives of this work package are:

Objective 4.1: Setup the project testbed, define interfaces used to provision hardware resources, and negotiate provisioning with the resource providers at the technical level.

Due to the isolation cloud technologies provide nowadays we do not need to setup a separate project testbed anymore. Instead, the EGI Federated Cloud resources available to the enmr.eu virtual organization (VO) are used, besides production, also to test the emerging solutions. Therefore also the provisioning contracts (SLAs) negotiated with EGI are applicable.

Objective 4.2: Define, implement, and deploy consolidated architecture for job submission and data access.

The architecture of job submission was defined in D4.1, and it is being gradually implemented and deployed, as described below. The architecture of data access was described in D4.4, and will be implemented in the following period (according to the schedule of T4.3).

Objective 4.3: Review existing security frameworks and define consolidated solution

The review and architecture definition was the subject of D4.2. Pilot implementation of the common IdP-SP proxy service is deployed and available, individual portals are being migrated to use it gradually.

Objective 4.4: Ensure smooth migration of the legacy portals to the consolidated architecture.

This work already started at the prototype and pilot level, and the objective is expected to be fully reached in the following period.

The work is organized in the following tasks:

- Task 4.1: Consolidation and operation of the infrastructure
- Task 4.2: Consolidation of job management mechanisms
- Task 4.3: Programmatic Access to datasets
- Task 4.4: Unified security and accounting model

Progress in each task is reported in the section below.

1.2.4.3 Work carried out in P1

Task 4.1: Consolidation and operation of the infrastructure

The inventory of the computing and storage resources supporting the project was completed for project milestone M4.1. These resources come mainly from the EGI HTC Platform and EGI Federated Cloud. 38 resource centres for a total of 211k CPU-cores are potentially accessible for use by West-Life project through the enmr.eu VO. In particular, seven of these resource centres: INFN-PADOVA (Italy), RAL-LCG2 (UK), TW-NCHC (Taiwan), SURFsara (The Netherlands), NCG-INGRID-PT (Portugal), NIKHEF (The Netherlands) and CESNET-MetaCloud (Czech Republic) have signed a SLA with EGI.eu which pledged 55 million hours of computing time and 61 TB storage capacity for the entire year 2016. Their further commitment for the entire year 2017 has been finalised in April 2017, with the addition of a new HTC provider from Spain, the IFCA-LCG2 site. In summary, in this new SLA eight resource centres have pledged 60 million hours of opportunistic grid computing time, 60 dedicated cores on the EGI Federated Cloud and more than 250 TB storage capacity to support the enmr.eu VO. Besides the committed resources, the first year of the project has seen the addition of new GPGPU HTC resources to the existing infrastructure available to West-Life: two resource centres, CIRMMP in Italy and Queen Mary University of London in UK, have made available respectively 3 nodes, each one with 12 CPU-cores and 2 NVIDIA Tesla K20m GPU cards, and 2 nodes: one with 32 CPU-core and 4 NVIDIA Tesla K80 GPU cards and one with 8 CPU-core and 1 NVIDIA Tesla K40c. These nodes could be included in the HTC e-infrastructure via the solution developed by INFN (in collaboration with EGI-Engage project) to enable GPGPU support in CREAM-CE, and allowed the AMPS-NMR (AMBER), DisVis and PowerFit web portals to greatly enhance their performance, as described in WP5 report. The inventory is kept updated on the internal project wiki at http://internal-wiki.west-life.eu/w/index.php?title=M11, including the resource usage.

While the production jobs of West-Life applications are mainly executed on the EGI HTC platform (or internal clusters local to the application portals), four resource centres that are part of the EGI Federated Cloud (CESNET-MetaCloud, INFN-PADOVA-STACK, IISAS-GPUCloud and IISAS-Nebula) are available for the enmr.eu VO. These have been used as pilot testbeds for rapid prototyping of innovative solutions like the ones described in documents M4.2 and D4.3, e.g. the use of DIRAC4EGI to submit jobs to cloud VMs, the deployment of GROMACS on the cloud via Cloudify orchestrator, and the use of INDIGO-DataCloud FutureGateway API for deploying on the cloud the AMPS-NMR portal. In this context, INFN partner has also procured at the beginning of 2017 a storage server with 24 TB of disk space installed at INFN-PADOVA-STACK cloud site with the INDIGO-DataCloud OneData software, to allow its testing with West-Life applications.

EGI HTC and Cloud platforms have been monitored and accounted by using the standard EGI tools. GStat monitoring tool has been decommissioned during the 2016 to be fully replaced by VAPOR, and EGI Accounting Portal has migrated to a renewed version at the beginning of 2017. The West-Life WP4 team has followed the evolutions of the EGI tools, often providing feedback and requirements for new features to EGI developers, and used VAPOR, ARGO and EGI Accounting Portal to obtain the needed information to control the status and report about the usage of the e-infrastructure accessible through the enmr.eu VO. The full report has been included in the deliverable document D4.3, timely delivered at project month 15. Figure 1.2.4.1 shows the numbers of HTC jobs submitted to gLite middleware directly compared to submissions via DIRAC4EGI. Figure 1.2.4.2 shows the corresponding success rate of those jobs; there is no clear dominance of reliability of either of the methods, therefore it is meaningful to preserve both ways with the existing portals as long as the gLite Workload Management System (WMS) is maintained.

Operations of the production infrastructure also means to handle issues with grid or cloud sites not properly configured or malfunctioning. Periodic test jobs are continuously submitted to resource centres supporting the enmr.eu VO through a Nagios server operated at INFN-PADOVA and badly working sites are informed via the EGI GGUS ticketing system. A total of 66 tickets have been handled during P1. INFN is also supporting via the Italian NGI the operations of enmr.eu VO central services like VOMS, LFC, and a number of gLite-WMS servers.

STFC has investigated the steps needed to join the EGI Federated Cloud as a Resource Provider of the SCD OpenStack cloud infrastructure. As a key component of a Resource Provider has been redesigned and redeveloped, the FedCloud integration at STFC has been postponed, however native interfaces have been made available for a proper workflow management of the virtual images used by W-L.

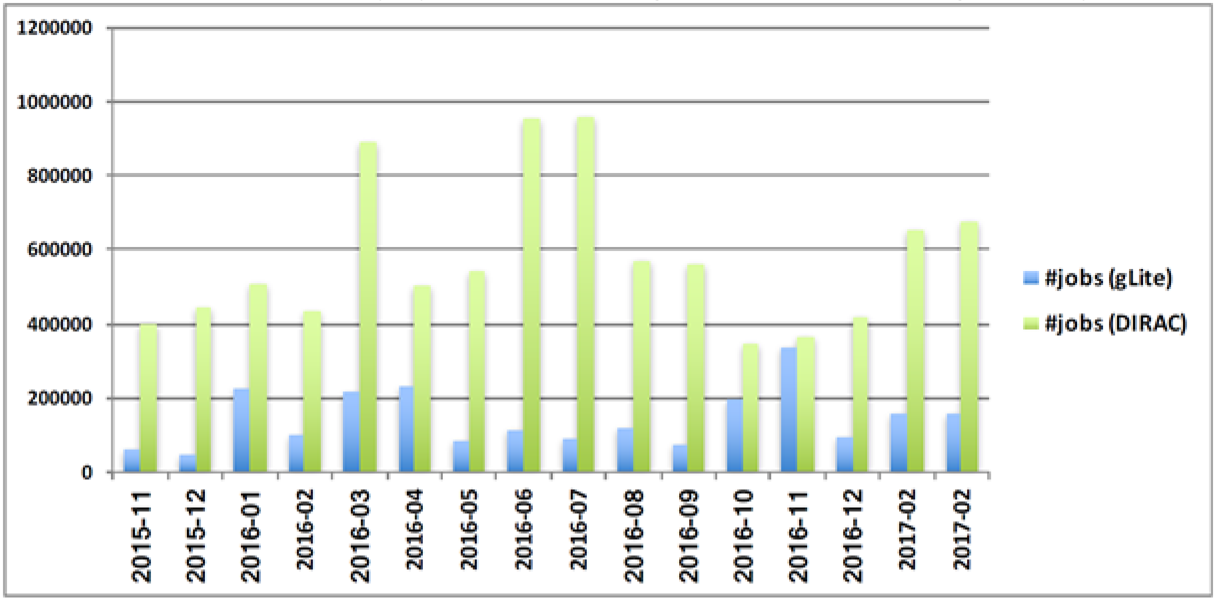


Figure 1.2.4.1: Numbers of jobs submitted via gLite and DIRAC

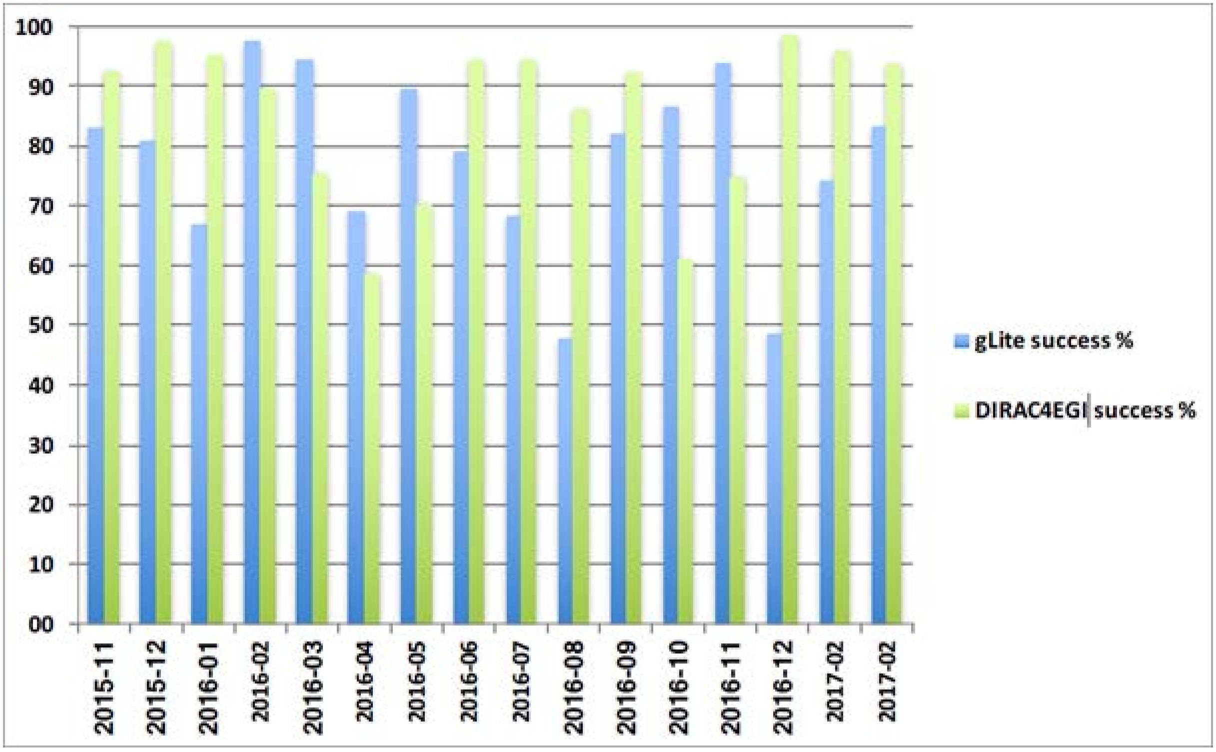


Figure 1.2.4.2: Success rate of jobs submitted via gLite and DIRAC

Task 4.2: Consolidation of job management mechanisms

Requirements of the involved scientific application portals as well as currently available technologies were analysed in detail in PY1 (reported in D4.1). The decision is to leverage the common layer of DIRAC for remote grid and cloud job submission, and to use cloud orchestration frameworks (with emphasis of H2020 INDIGO DataCloud project solutions) for deployment of new portal versions in the EGI Federated Cloud.

The HADDOCK portal operated by UU, which is responsible for the large majority of jobs on grid resources uses DIRAC4EGI for submission. CCP4Online has recently been extended to allow submission of grid jobs through the enmr.eu VO using DIRAC4EGI. For the new GPGPU-enabled portals (DisVis and Powerfit), gLite direct submission to the GPGPU-enabled CEs was implemented since those directly target the compute elements and do not use the WMS brokering system. Migration of the Gromacs portal to the new version of the application software (MU) was already done using the cloud orchestration approach. MU and CSIC analysed the technical requirements of Scipion and Scipion Web Tools applications in order to integrate them on the proposed architecture, a prototype was developed.

In collaboration with WP2, training material on cloud orchestration of scientific application portals was prepared, and a training event with participation of several project partners was organized (details in WP2 report).

Task 4.3: Programmatic Access to datasets

This task started in PY2 according to the original workplan. The first achievement is the availability of the virtual folder, a mechanism to unify access to data from various sources in the web portals. The virtual folder is available in two ways. A VM provides programmatic access to datasets by file system mounts. In addition, it has been integrated into the West-Life portal. This acts as a proof of concept for integration into partner web services too. A review of data sources has shown the importance of also integrating access to data stored at experimental facilities.

Distribution of the virtual folder and associated software dependencies is done using CernVM-FS which has shown particular benefits on grid or cloud deployments especially when connected with caching proxy. First metadata service API was established within WP6 virtual folder to define inputs of dataset entries and the UI components has been established with EBI component library to collect dataset information and to allow refine/browse existing entries published in public EBI database. This work is ongoing.

Review of programmatic access to various data sources was done in D4.4, and the findings will be used in further integration of the portals with those data sources. A major challenge in this work is the size of the primary data for cryoEM facilities, probably the largest datasets involved in the whole work of the project. The right approach is currently discussed among project partners with direct experience in the field (CSIC, STFC, and MU to some extent).

Task 4.4: Unified security and accounting model

During the reported period we focused on developing a unified security mechanism that portal applications in the project could use for authentication and controlling access of their users (AAI). The AAI design is described in the project deliverable D4.2 “Common security model design”. Building on current best practices and development in e-Infrastructures and similar research environments, the new security model harmonizes existing approaches to AAI and provides a common solution for all West-Life systems. The technology allows for a gradual transition to new services so that we can easily move existing services to new AAI, and will support integration with ELIXIR’s planned AAI system.

The new mechanism is based on the SAML protocol and utilizes existing identity federations. The concept of IdP-SP-Proxy has been introduced as the main component of the new AAI architecture. The proxy makes it possible to unify access to multiple identity providers and centralize handling of users' identifiers and attributes.

In order to demonstrate the viability of the concepts, we have deployed an instance of the IdP-SP-Proxy service. The Proxy was configured to support users' authentication to several identity providers. At the moment, it supports ARIA, Google, ORCID, and we added authentication using the existing WeNMR SSO system and passwords.

On the level of end-service, we established a simple web application that demonstrates how SAML-based security is integrated with the services. We also provided an adaptation of the GROMACS portal that implements a basic SAML-based authentication to an existing application. Specific plans to integrate the unified security model with Powerfit, DisVis and Scipion portals (in the first batch) were drawn, and the code is being adapted.

1.2.4.4 Progress towards Deliverables and Milestones

Number	Title	Lead	Type /planned delivery date	Progress
D4.1	D4.1: Consolidated architecture of job submission and interaction with infrastructure	Luna	Public report Month 8	The deliverable concludes the review and design work in the first months of the project. D4.1 was delivered on time.
D4.2	D4.2: Common security model design	MU	Public report Month 9	This is a complementary deliverable to D4.1, specifically focused on the security aspects. Otherwise the preparation of D4.1 and D4.2 was closely coordinated. D4.2 was delivered on time.
D4.3	D4.3: Report on experience with deployment of consolidated platform and its interaction with infrastructure	INFN	Public report Month 15	This report summarizes the achievements and experience of the first deployment (Milestone M4.2). D4.3 was delivered on time.
D4.4	D4.4: Overview of external datasets, strategy of access methods and implementations on the portal architecture	STFC	Public report Month 18	Plans for access to experimental data and processed results. D4.4 was delivered on time.
D4.5	Report on progress of the deployment of consolidated platform and its interaction with infrastructure (month 26)	INFN	Public report Month 26	
D4.6	Final report on deployment of consolidated platform and the overall architecture	MU	Public report Month 36	

Milestone Number	Milestone Name	Estimated Date	Means of Verification
MS11	Inventory of available resources and testbed setup	4 / Delivered	Documentation on wiki
MS12	First deployment of the consolidated platform	12 / Delivered	The architecture defined in D4.1 is deployed in production and API specifications published in the wiki
MS13	Prototype access to selected datasets	Month 22	Reporting Period 2 Datasets available for relevant portal applications Partly achieved already
MS4	Second deployment of the consolidated platform	Month 24	Reporting Period 2 New interfaces/services integrated in the platform ready for production use by other WPs and published in the wiki
MS15	Final deployment of the consolidated platform	Month 34	New interfaces/services integrated in the platform ready for production use by other WPs and published in the wiki

1.2.4.5 Deviations from Annex 1

At the technical level, there are no deviations from the Annex.

The work on design and prototyping of the consolidated job management mechanisms (cloud orchestration in particular) was concentrated more in P1, yielding more effort to be claimed with WP4 especially by MU. This was compensated by spending less effort with WP7. We expect the opposite ratio in the following period, ending up with the total effort claimed as originally planned.

1.2.5 Work package: 5 Support

1.2.5.1 Summary and Significant Results

In this WP centered on building and operating the Virtual Research environment, we set up the web front end of West-Life using the ARIA content management system (CMS) provided and maintained by Instruct. Content including documentation and training material was added and migrated to this new site. A survey about current means of user support and how it can be added to West-Life was conducted. Furthermore, a first version of the West-Life portal was launched.

Two major shortcomings due to missing crosstalk of e-science portals were selected to be tackled by the VRE including *data management*, with users having to manually download their files on their laptop and then re-upload them on the e-science portal they want to use, and *authentication*, with users having to manage multiple sets of credentials for different e-science portals.

It was thus decided to provide a virtual folder as a self-contained appliance, that portals can use as a storage backend (a unified mount point). The objective is to provide users with the capability of browsing all their files with a consolidated view, and choose which file to use on the e-science portal. The file is then directly sent from the storage location (ex: EUDAT) to the e-science portal server. In addition, we are aiming at a single method of user authentication to each e-science portal by interconnecting with the ARIA SSO (which does not have all the capabilities people expect, but is already widely used and is expected to last).

New services were created including a first version of ScipionCloud virtual appliance as well as two new web portals, DisVis and PowerFit, utilizing GPGPU resources on the EGI grid. In addition, some new workflows were created establishing direct links between existing service portals.

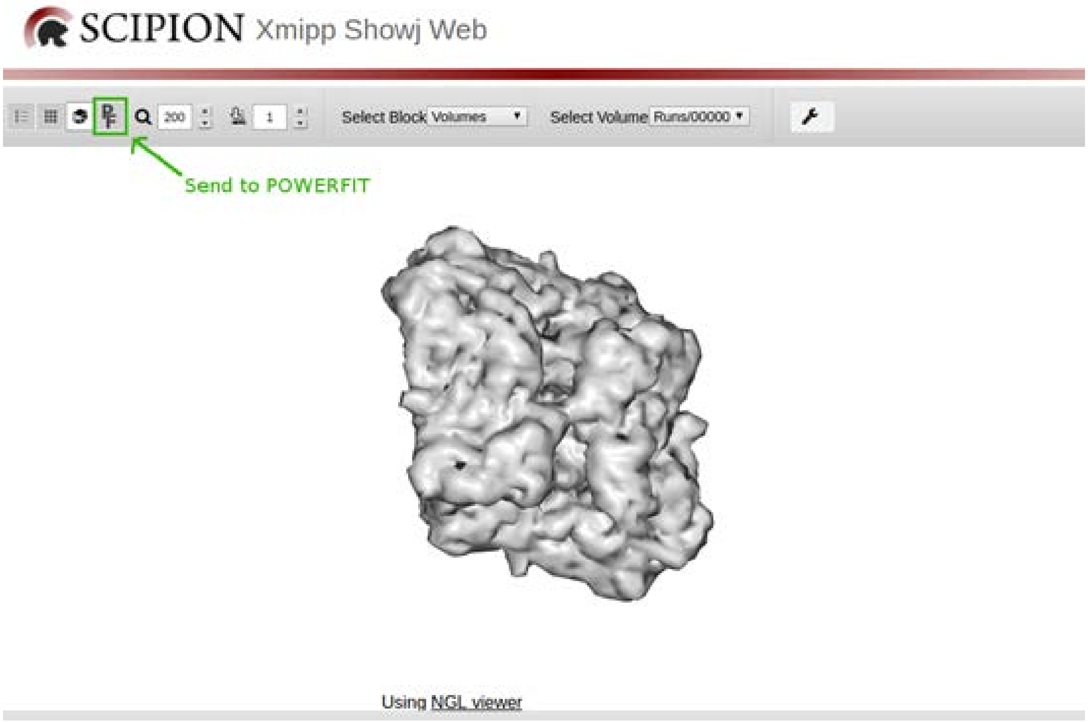


Figure 1.2.5.1: The Scipion EM tool can now submit directly to the PowerFit web portal

1.2.5.2 Objectives and Tasks

The objective of this WP is to integrate existing and operating services from the WeNMR Virtual Research Community and West-Life partners into a Virtual Research Environment (Objective 5.1). Within the scope of this research environment relevant information, training material, as well as user support are provided to engage and support the community (Objectives 5.2 and 5.3). Easy distribution and maintenance of services on local infrastructures or on EGI federated cloud resources is intended to be facilitated by customized end-user VMs and/or application containers (e.g. Docker), built for various scenarios (Objective 5.4). And finally, new web interfaces for relevant software solutions will be added to the VRE to increase its versatility and to make it relevant for a broader audience (Objective 5.5).

To achieve those objectives, the following tasks were defined:

Task 5.1 – Deployment and operation of the West-Life portal (Luna, all)

The impact of the West-Life VRE is intended to be measured by monitoring the usage of services across all partners together with those of the WeNMR VRC after defining a baseline at the beginning of the project. CryoEM services are provided with the SCIPION platform for software integration. All services are gathered in a new VRE portal which provides end users with a friendly and dynamical entry point to the services themselves. It furthermore acts as a knowledge and support center. The portal is built on innovative technology developed by LUNA and we aim to migrate when possible existing portals to make direct use of the technology solutions offered by LUNA. Another important task is the harmonization of user authentication and authorization mechanisms (AAI). The choice and implementation of AAI mechanism is done in close collaboration with EGI- Engage to maximize compatibility and impact. The new VRE portal also implements tools and services related to data discovery and access (see WP6).

Task 5.2 – Knowledge and support center (Instruct, all)

Building on the existing knowledge and support center of WeNMR, covering NMR and SAXS services we migrate existing content 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. As in Task 5.1, this is being done be done in close collaboration with the related EGI-Engage Competence Centers to minimize heterogeneity and maximize impact.

Task 5.3 – Development and integration of new service portals (UU, all)

This task aims at integrating EGI Grid infrastructure resources into existing portals and services with limited local resources targeting the most suited e-Infrastructure solution(s), be it Grid, Cloud or HPC resources. Again, the interaction with various Competence Centers under the EGI-Engage project, especially the MoBrain Competence Center, to which several partners of West-Life VRE participate (UU, CSIC, CIRMMP, INFN and STFC) is a major contributing factor. Care is taken to offer user-friendly interfaces, by choosing the most suited submission mechanisms (e.g. the efficient DIRAC4EGI service), offering AAI single sign-on (SSO) capabilities and shared data management among services provided by West-Life.

Task 5.4 – Customized end-users VMs (STFC, all)

To truly exploit structural biology data a wide range of techniques and expertise is often required. In most cases, users however are expert in only one or a few techniques and associated software. In this task, we build custom VMs for different use cases, including all the necessary software, documentation and examples for complex workflows. These VMs are 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) have been provided, to allow use on both the EGI Federated Cloud and local OpenStack/Nebula cloud resources for example. Depending on the requirements of the VM, local installation on a user's laptop is also feasible. The use of these containers also provides an attractive mechanism for offering services where external use of the West-Life services is not an option e.g. in a commercial setting in companies.

1.2.5.3 Work Carried Out In P1

Task 5.1 – Deployment and operation of the West-Life portal (Luna, all)

As part of measuring the impact of the services offered by West-Life, we are constantly monitoring their usage. Their baseline was defined based on statistics collected over 2015 (i.e. including a 10 months period prior to the start of the project) and reported in D5.2. In this report we compared the baseline with the 2016 statistics, which allows us to monitor the KPIs defined for our portals. The number of users, job submissions (to the portal) and grid job submission (by the portal if applicable) are listed in Table 1.2.5.1.

Table 1.2.5.1: 2016 Portal statistics for services offered by the West-Life partners

	Users total	***Users 2016***	***Job submissions 2016***	***Grid/cloud jobs 2016***
AMPSNMR	390	90	185	3655
ARP/wARP	4418	666	3466	0
AutoRickshaw	2550	430	4357	0
CCD	0	0	4000	0
CCP4-AMPLE*	187	187	535	0
CCP4-BALBES	1521	700	3148	0
CCP4-Crank2*	94	94	587	0
CCP4-MORDA*	202	202	790	0
CCP4-MRBUMP	874	453	1721	0
CCP4-SHELX*	151	151	656	0
CCP4-ZANUDA	390	196	598	0
CS-ROSETTA3	59	9	33	263871
DISVIS*	39	38	172	64
GROMACS	121	9	107	368
HADDOCK	8320	1621	27291	8000000
PDB_REDO	1200	500	4400	0
POWERFIT*	21	21	34	13

SCIPION	-	-	418	30446
UNIO	80	21	72	280
VICI	417	131	196	0

* - services added in 2016

From these statistics it should be clear that all portals are being used, with some generating a very large number of grid jobs. A comparison with the defined baseline (D5.2) allows us to monitor their usage in terms of growth or decrease in number of users and job submission compared to the previous reporting period (Table 1.2.5.2). Note that the newly added portals are not listed in Table 1.2.5.2 since no baseline was defined for those.

Table 1.2.5.2: KPIs of West-Life services showing change in usage compared to the baseline defined in D5.2

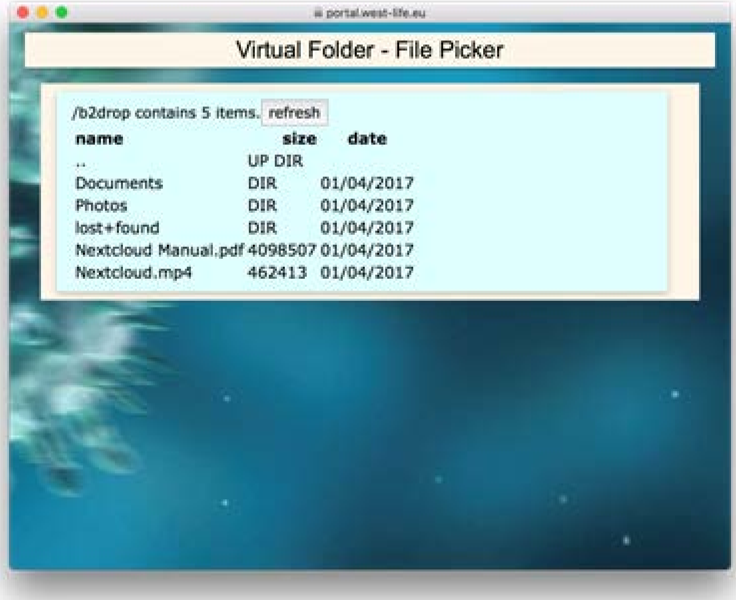
	%increase #Users total	%increase #Users 2016	%increase #Job submissions	%increase #Grid/cloud jobs
AMPSNMR	30%	80%	n.a.	-54%
ARP/wARP	8%	57%	7%	n.a.
AutoRickshaw	10%	-6%	22%	n.a.
CCD	n.a.	n.a.	33%	n.a.
CCP4-BALBES	32%	-4%	-5%	n.a.
CCP4-MRBUMP	51%	-4%	32%	n.a.
CCP4-ZANUDA	47%	4%	61%	n.a.
CS-ROSETTA3	16%	-44%	-51%	40%
GROMACS	8%	-65%	-38%	-37%
HADDOCK	24%	12%	10%	7%
PDB-REDO	71%	0%	47%	n.a.
SCIPION	n.a.	n.a.	318%	1347%
UNIO	36%	-32%	16%	-88%
VICI	43%	42%	111%	n.a.

Compared to the defined baseline (based on the 2015 statistics) most portals show a sustained growth while reduction in usage is observed only in a few portals (Gromacs, CS-Rosetta3). This can be explained by somewhat outdated software versions. Within the West-Life project these services are being updated including the software as well as the front end in order to meet user requirements. Such KPIs also provide a means to prioritize our efforts and could even lead to the closure of a portal if its usage will completely drop. In that respect, the portfolio of services offered by West-Life should be considered as dynamical to reflect usage and user needs.

Next to monitoring and adding new services we also tested some integration concepts for the West-Life portal and existing e-science portals. I.e. could all the portals, which are effectively managed and running at various locations, be gathered in a simple manner within one interface (the current solution is redirecting to the various locations). We tested crawling technologies (so that the main portal could replicate existing West-Life portals' job submit pages), but had to acknowledge that this approach was not universally suited due to the different web technologies in use and different requirements on user input among West-Life e-science portals. It became apparent that the constraints imposed by this approach on e-science portal operators was not compatible with the diverse services offered. Unifying all those portals would require many more PM efforts than is available within the project. It was therefore decided to use a more lightweight approach with focus on 2 key points for convergence: SSO and data management.

We therefore set up a basic web portal as a single point of entry to all West-Life portals providing documentation, training, and links to all e-science portals and interconnected this portal with Instruct ARIA SSO.

For data management, we first implemented a web interface that consolidates all the user data from Dropbox, Google Drive, Amazon, and EUDAT. We then replaced those connectors by the virtual folder developed by STFC and reported in Deliverable 6.1, so that the West-Life portal could be a testbed and a blueprint for other West-Life associated portals (interconnect with SSO + virtual folder).



/b2drop contains 5 items.

name	size	date
..	UP DIR	
Documents	DIR	01/04/2017
Photos	DIR	01/04/2017
lost+found	DIR	01/04/2017
Nextcloud Manual.pdf	4098507	01/04/2017
Nextcloud.mp4	462413	01/04/2017

Figure 1.2.5.2: File Picker demo for navigating to files in a connected B2Drop account

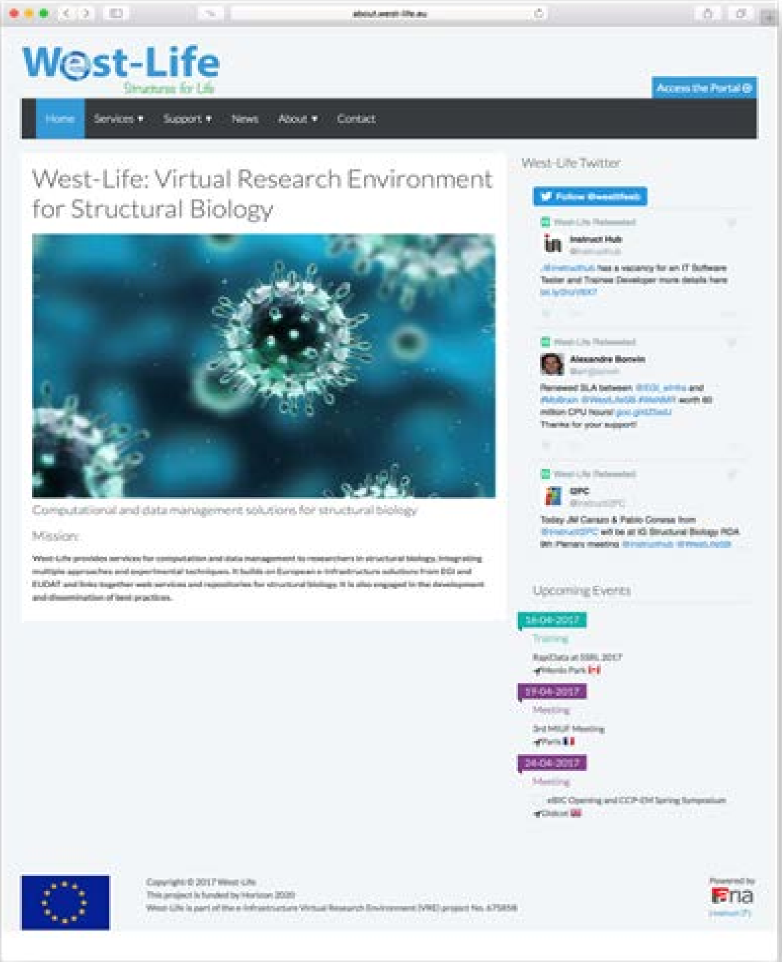
The ultimate objective is to have all portals sharing ARIA SSO and virtual folder capabilities, and share the main West-Life portal as single point of entry for documentation etc. Concretely, to minimize overhead for existing West-Life portals, we are working toward integrating the ARIA SSO, providing support and guidance to the various portal operators and developers in doing that. We furthermore provide a web widget to leverage the virtual folder capabilities. This widget will enable end users to send one of their files (be it stored in EUDAT, Dropbox, etc.) directly to the portal, instead of manually uploading a file from their laptop.

Task 5.2 – Knowledge and support center (Instruct, all)

During the July 2016 developers meeting (Paris) new documentation was added to a West-Life WordPress site as well as existing documentation migrated from the WeNMR portal. This documentation has now been migrated to the main portal (under ARIA CMS – see below), which will increase the long-term sustainability of the web front end since it is part of the machinery managed by INSTRUCT.

Documentation about how to exploit GPGPUs in both HTC and Cloud EGI platforms has been added in 2017, and is publicly available at the main portal under the menu Support/Documentation/EGI Platforms/Accelerated Computing Platforms.

For grid/cloud related issues, the main support mechanism is via the EGI GGUS ticketing service (https://ggus.eu). A total of 66 GGUS tickets related to West-Life/WeNMR have been handled during P1.



West-Life: Virtual Research Environment for Structural Biology

Computational and data management solutions for structural biology

Mission:

West-Life provides services for computation and data management to researchers in structural biology, integrating multiple approaches and experimental techniques. It builds on European e-Infrastructure solutions from EGI and EUDAT and links together web services and repositories for structural biology. It is also engaged in the development and dissemination of best practices.

16-04-2017
Training
RapiData at SSRL 2017
Menlo Park

19-04-2017
Meeting
3rd MIUF Meeting
Paris

24-04-2017
Meeting
eBIC Opening and CCP-EM Spring Symposium
Didcot

Copyright © 2017 West-Life
This project is funded by Horizon 2020
West-Life is part of the e-Infrastructure Virtual Research Environment (VRE) project No. 675858

Powered by
aria

Figure 1.2.5.3: West-Life web presence

In February 2017, the executive steering committee decided on using the ARIA content management system (CMS) as the platform for the West-Life web presence. Features of this system supporting the decision include:

- A customizable front-end with menus and submenus which also provides a link to the portal
- The system can dynamically pull news items/events/jobs from the Instruct implementation of ARIA (Instruct aggregates a large body of current information from the structural biology

community worldwide)
- ARIA CMS can hold the support documentation pages for the tools implemented in the VRE, maintaining the integrity of the interface
- The system makes it easy to edit the documentation and enable authorized users to add/edit/delete sections of or delete webpages (This can be done using a WYSIWG text editor, Instruct furthermore hosts workshops where administrators can get trained in using the system)
- The homepage can include a twitter feed as in the current homepage
- ARIA CMS has an integrated forum for discussion and support activities
- ARIA authentication is plugged into IdP services avoiding the need for registration to other services
- Websites hosted on the ARIA platform have multiple moderators/administrators who need not be related to Instruct
- ARIA is a cloud service managed by Instruct with services being used by different organizations, thus ensuring continued service life of the website

The portal integrates a direct Twitter feed from the West-Life Twitter channel (@westlifesb) and a list of upcoming events (some of which inherited directly from the INSTRUCT web site when relevant) (Figure 1.2.5.4)

West-Life Retweeted
Instruct Hub
@instructhub
.@instructhub has a vacancy for an IT Software Tester and Trainee Developer more details here bit.ly/2nzV8X7

West-Life Retweeted
Alexandre Bonvin
@amjbonvin
Renewed SLA between @EGI_eInfra and #MoBrain @WestLifeSB #WeNMR worth 60 million CPU hours! goo.gl/dZ5xdJ
Thanks for your support!

16-04-2017
Training
RapiData at SSRL 2017
Menlo Park

19-04-2017
Meeting
3rd MIUF Meeting
Paris

24-04-2017
Meeting
eBIC Opening and CCP-EM Spring Symposium
Didcot

Figure 1.2.5.4: News and Twitter Feed integration

The ARIA CMS has so far been implemented and is accessible at west-life.eu. Documentation from the WordPress site has been migrated and content is continuously being added.

In order to streamline user support and in preparation for Deliverable D5.4, a list of current support mechanisms and how to best incorporate them into the West-Life forum has been gathered:

Table 1.2.5.3: Overview of current support mechanisms of West-Life services and their envisaged integration into the West-Life portal

Portal	***Current Support form***	***Preferred Integration with West-Life support forum***
Scipion	Public archive mailing list (https://lists.sourceforge.net/lists/listinfo/scipion-users)	Link from West-Life forum
GROMACS	ask.bioexcel.eu	Link from West-Life forum
ViCi	On http://www.embl-hamburg.de/vici there are brief introduction to the service and instructions on its use. There is also a link for a direct email contact to the developers.	Link from West-Life forum
HADDOCK	ask.bioexcel.eu	Link from West-Life forum
DisVis	ask.bioexcel.eu	Link from West-Life forum
PowerFit	ask.bioexcel.eu	Link from West-Life forum
AMPS-NMR	http://www.wenmr.eu/wenmr/tutorials/nmr-tutorials/amber	Port to West-Life forum
CS-Rosetta3	https://www.wenmr.eu/wenmr/help-center/nmr-services-issues/cs-rosetta	Port to West-Life forum
FANTEN	http://fanten-enmr.cerm.unifi.it:8080/manual/FANTEN_manual.pdf	Link from West-Life forum
UNIO	https://www.wenmr.eu/wenmr/help-center/nmr-services-issues/unio	Port to West-Life forum
XPLOR-NIH	http://www.wenmr.eu/wenmr/tutorials/nmr-tutorials/xplor-nih	Port to West-Life forum

ARP/wARP	On http://www.embl-hamburg.de/ARP/ there are links to the manual, FAQ and tutorials. There is also a link for a direct email support by the developers.	Link from West-Life forum
Auto-Rickshaw	On http://www.embl-hamburg.de/Auto-Rickshaw/intro.htm there are links to the introduction, FAQ and a link for a direct email support by the developers.	Link from West-Life forum
CCP4 - Ample** ***CCP4 - Balbes*** ***CCP4 - Crank2***** ***CCP4-MoRDa**** ***CCP4 - MrBUMP*** ***CCP4 - Shelx***** ***CCP4 - Zanuda***	Email distribution list serving CCP4 community (searchable, but sign-up required to post): http://www.jiscmail.ac.uk/lists/ccp4bb.html There have been periodic suggestions to move to forum-based system, but all fail to get support from community. There is also an email helpdesk for specific queries about CCP4 software and web services ccp4@ccp4.ac.uk The courses page on the CCP4 website provides access to recent talks and tutorials on the CCP4 suite.	Link from West-Life forum
PDB-REDO	Direct email to developers (pdbredo@gmail.com) and CCP4 bulletin board + proactive support by the developers (i.e. users with apparent problems are contacted before a support request is made).	Link from West-Life forum
CCD	Direct email to developers (proteinccd@gmail.com)	Link from West-Life forum
PDBe API	Mailing list - pdbe-api-users@ebi.ac.uk	

Figure 1.2.5.5: Integrated and linked West-Life forums

To guide users new Tutorials were created (3 x HADDOCK, 2 x PowerFit, 1 x DisVis, 3 x Scipion) which are accessible from the West-Life site: http://about.west-life.eu/network/west-life/support/tutorials

Additionally, in coordination with BioExcel, webinars for West-Life services including DisVis, PowerFit, and HADDOCK were held and are available to the community as well at: http://about.west-life.eu/network/west-life/support/webinars
Also lectures and tutorials of the Cryo-EM Workshop held at the International Symposium on Grids and Clouds (ISGC, 2017 Taipei) have been recorded and will be made available to the community on YouTube and the West-Life site.

Robust solutions for cryoEM fitting and visualisation of interaction space

BioExcel Webinar Series

Presenters: Gydo van Zundert
with Mikael Trellet and Jörg Schaarschmidt
Host: Adam Carter

15 February, 2017
15:00 GMT / 16:00 CET

Webinar: Robust solutions for cryoEM fitting and visualisation of interaction space

41 views

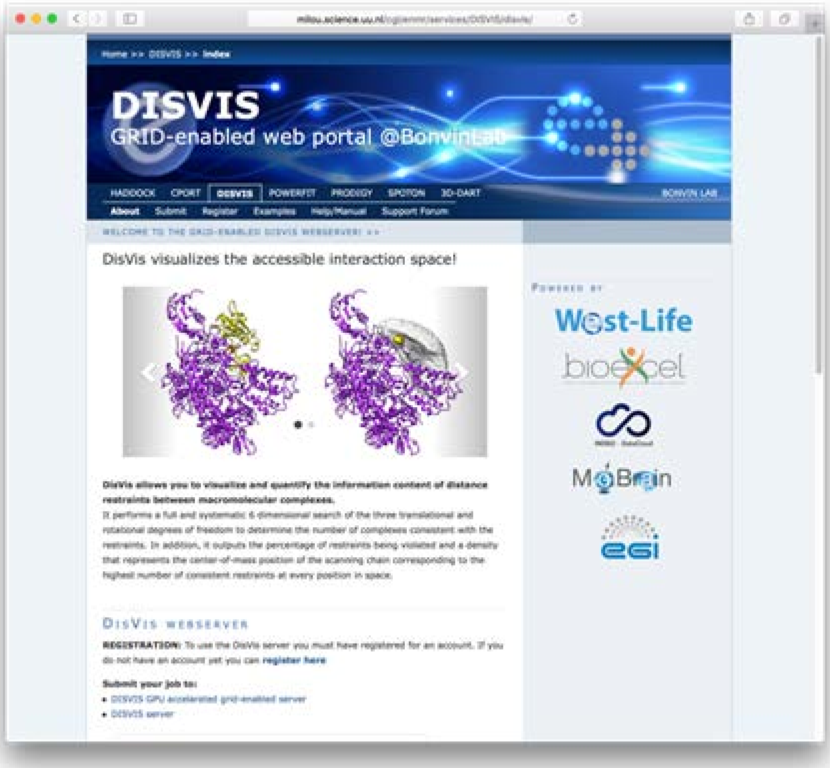
BioExcel Webinar Series #7: PRODIGY, a web server to predict binding affinities in protein

Figure 1.2.5.6: Webinar for West-Life services

Task 5.3 – Development and integration of new service portals (UU, all)

Two new web portals using grid GPGPU resources were developed and put in production covering workflows addressing structural fitting into cryoEM data (PowerFit) as well as analysis of mass spectrometric data (DisVis).

DisVis visualizes and quantifies the information content of distance restraints between macromolecular complexes. It performs a full and systematic 6-dimensional search of the three translational and rotational degrees of freedom to determine the number of complexes consistent with the restraints. In addition, it outputs the percentage of restraints being violated and a density that represents the center-of-mass position of the scanning chain corresponding to the highest number of consistent restraints at every position in space.



DisVis visualizes the accessible interaction space!

DisVis allows you to visualize and quantify the information content of distance restraints between macromolecular complexes.
It performs a full and systematic 6-dimensional search of the three translational and rotational degrees of freedom to determine the number of complexes consistent with the restraints. In addition, it outputs the percentage of restraints being violated and a density that represents the center-of-mass position of the scanning chain corresponding to the highest number of consistent restraints at every position in space.

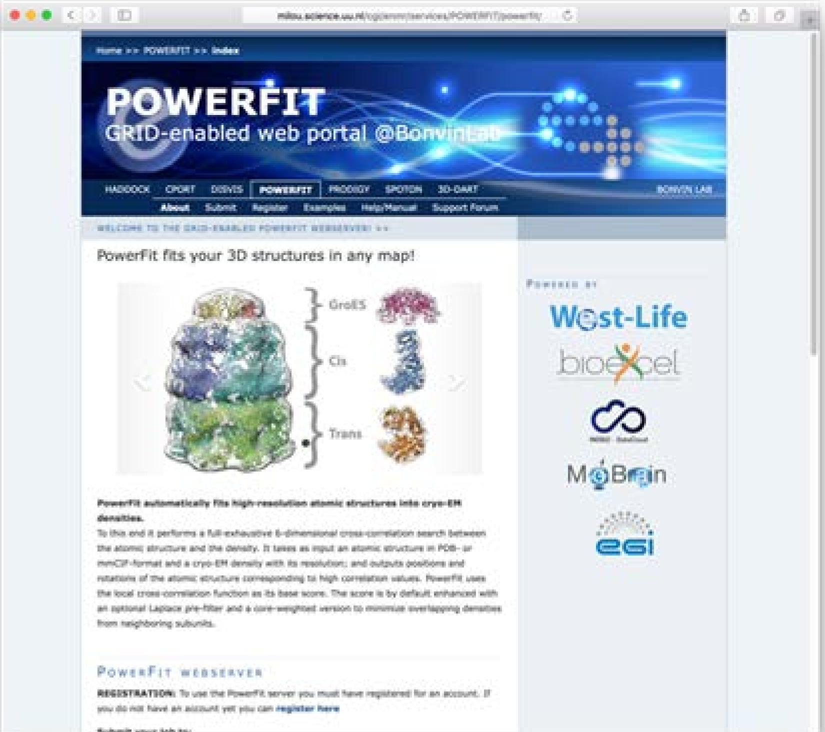
DisVis webserver

REGISTRATION: To use the DisVis server you must have registered for an account. If you do not have an account yet you can **register here**

Submit your job to:
- DISVIS GPU accelerated grid-enabled server
- DISVIS server

Figure 1.2.5.7: DisVis webserver landing page

PowerFit automatically fits high-resolution atomic structures into cryo-EM densities. To this end it performs a full-exhaustive 6-dimensional cross-correlation search between the atomic structure and the density. It takes as input an atomic structure in PDB- or mmCIF-format and a cryo-EM density with its resolution and outputs positions and rotations of the atomic structure corresponding to high correlation values. PowerFit uses the local cross-correlation function as its base score. The score is by default enhanced with an optional Laplace pre-filter and a core-weighted version to minimize overlapping densities from neighbouring subunits.



PowerFit fits your 3D structures in any map!

PowerFit automatically fits high-resolution atomic structures into cryo-EM densities.
To this end it performs a full-exhaustive 6-dimensional cross-correlation search between the atomic structure and the density. It takes as input an atomic structure in PDB- or mmCIF-format and a cryo-EM density with its resolution; and outputs positions and rotations of the atomic structure corresponding to high correlation values. PowerFit uses the local cross-correlation function as its base score. The score is by default enhanced with an optional Laplace pre-filter and a core-weighted version to minimize overlapping densities from neighboring subunits.

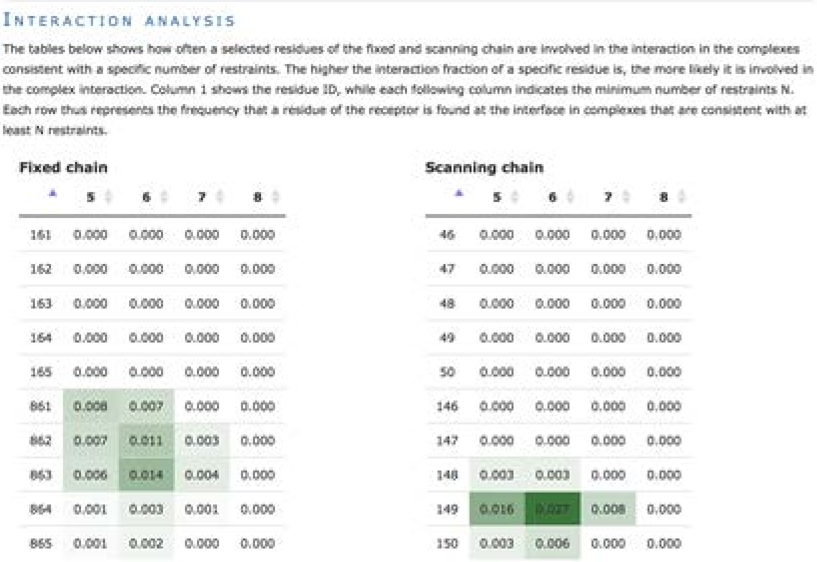
POWERFIT WEBSERVER

REGISTRATION: To use the PowerFit server you must have registered for an account. If you do not have an account yet you can **register here**

Submit your job to:

Figure 1.2.5.8:PowerFit webserver landing page

Next to providing a simple frontend to the applications and access to computing resources for performing the actual calculations, both portals are a valuable source of information for users by featuring help pages, example in- and output, as well as links to the Tutorials, forums, Webinars. Furthermore, they feature rich result pages which do not merely provide access to the generated results but assist the user in interpreting these by providing explanations, highlighting relevant entries and displaying auto generated images of the structural output, where possible.



INTERACTION ANALYSIS

The tables below shows how often a selected residues of the fixed and scanning chain are involved in the interaction in the complexes consistent with a specific number of restraints. The higher the interaction fraction of a specific residue is, the more likely it is involved in the complex interaction. Column 1 shows the residue ID, while each following column indicates the minimum number of restraints N. Each row thus represents the frequency that a residue of the receptor is found at the interface in complexes that are consistent with at least N restraints.

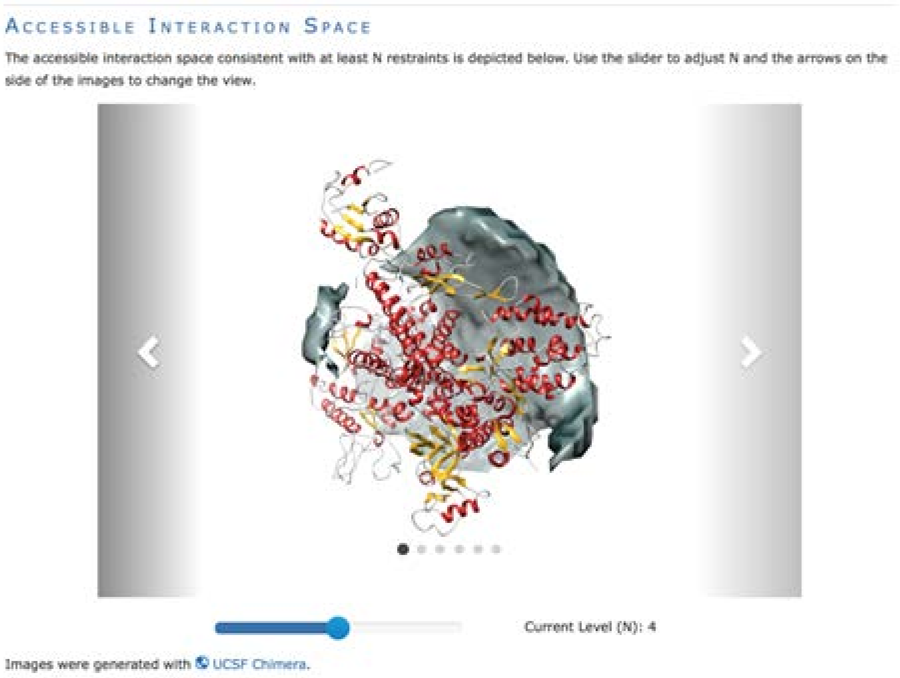
Fixed chain

▲	5	6	7	8
161	0.000	0.000	0.000	0.000
162	0.000	0.000	0.000	0.000
163	0.000	0.000	0.000	0.000
164	0.000	0.000	0.000	0.000
165	0.000	0.000	0.000	0.000
861	0.008	0.007	0.000	0.000
862	0.007	0.011	0.003	0.000
863	0.006	0.014	0.004	0.000
864	0.001	0.003	0.001	0.000
865	0.001	0.002	0.000	0.000

Scanning chain

▲	5	6	7	8
46	0.000	0.000	0.000	0.000
47	0.000	0.000	0.000	0.000
48	0.000	0.000	0.000	0.000
49	0.000	0.000	0.000	0.000
50	0.000	0.000	0.000	0.000
146	0.000	0.000	0.000	0.000
147	0.000	0.000	0.000	0.000
148	0.003	0.003	0.000	0.000
149	0.016	0.027	0.008	0.000
150	0.003	0.006	0.000	0.000

Figure 1.2.5.9: Rich result pages of the new portals,
which provide explanations and highlight important data



ACCESSIBLE INTERACTION SPACE

The accessible interaction space consistent with at least N restraints is depicted below. Use the slider to adjust N and the arrows on the side of the images to change the view.

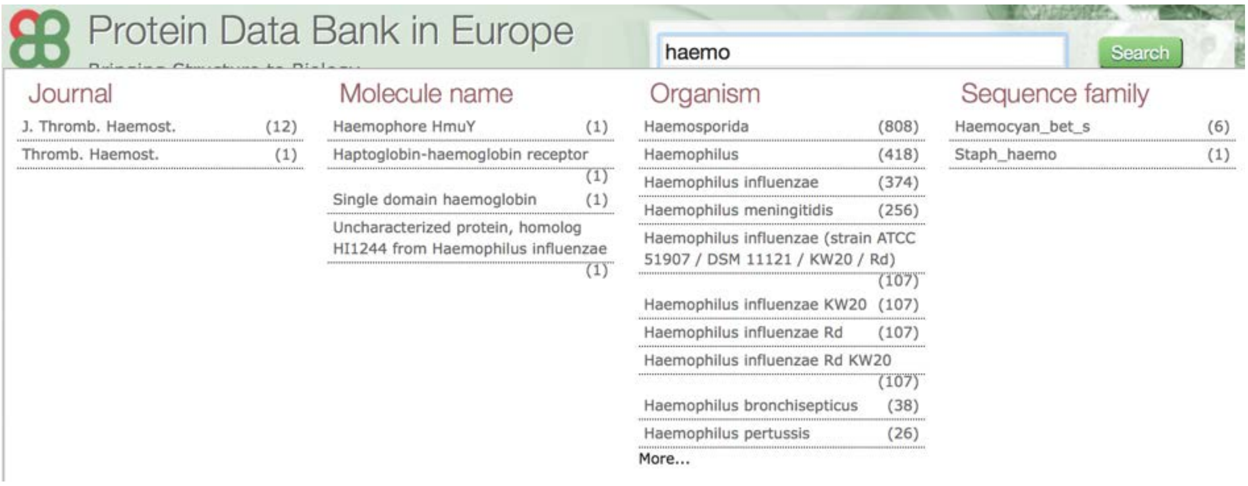
Current Level (N): 4

Images were generated with UCSF Chimera.

Figure 1.2.5.10: auto generated images provide a first visual impression of the results

In the case of PowerFit, the newly added portal is even part of a new workflow enabling users to directly fit structures into cryo-EM maps generated with Scipion (see Figure 1.2.5.1 and WP7 for further details).

CCP4Online provides online implementations of selected pipelines from the full CCP4 suite, and the range of pipelines made available is continually being expanded. Ample, Crank2 and Shelx portals were added by CCP4 at the beginning of 2016, and were included in the document of baseline services but with no available usage statistics. A further CCP4 service MoRDa was added in July 2016, and was thus not included in the baseline document. MoRDa is an alternative to MrBUMP or Balbes for crystallographic structure solution by molecular replacement. It has its own domain database, and models relevant to the target sequence are further adjusted before the molecular replacement search. These new CCP4 portals have been developed using the existing CCP4 framework for online services. Work is ongoing to provide REST APIs and integration with the Virtual Folder for all CCP4 online portals.
Several PDBe web components, have been developed and been integrated into the Virtual Folder via utilization of the PDBe REST API as reported in D5.5. Within the workflow, the Virtual Folder page allows users to search the PDBe database using the PDBe autocomplete feature (Figure 1.2.5.11) provided by the PDBe Search API.



Journal	
J. Thromb. Haemost.	(12)
Thromb. Haemost.	(1)

Molecule name	
Haemophore HmuY	(1)
Haptoglobin-haemoglobin receptor	(1)
Single domain haemoglobin	(1)
Uncharacterized protein, homolog HI1244 from Haemophilus influenzae	(1)

Organism	
Haemosporida	(808)
Haemophilus	(418)
Haemophilus influenzae	(374)
Haemophilus meningitidis	(256)
Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd)	(107)
Haemophilus influenzae KW20	(107)
Haemophilus influenzae Rd	(107)
Haemophilus influenzae Rd KW20	(107)
Haemophilus bronchisepticus	(38)
Haemophilus pertussis	(26)

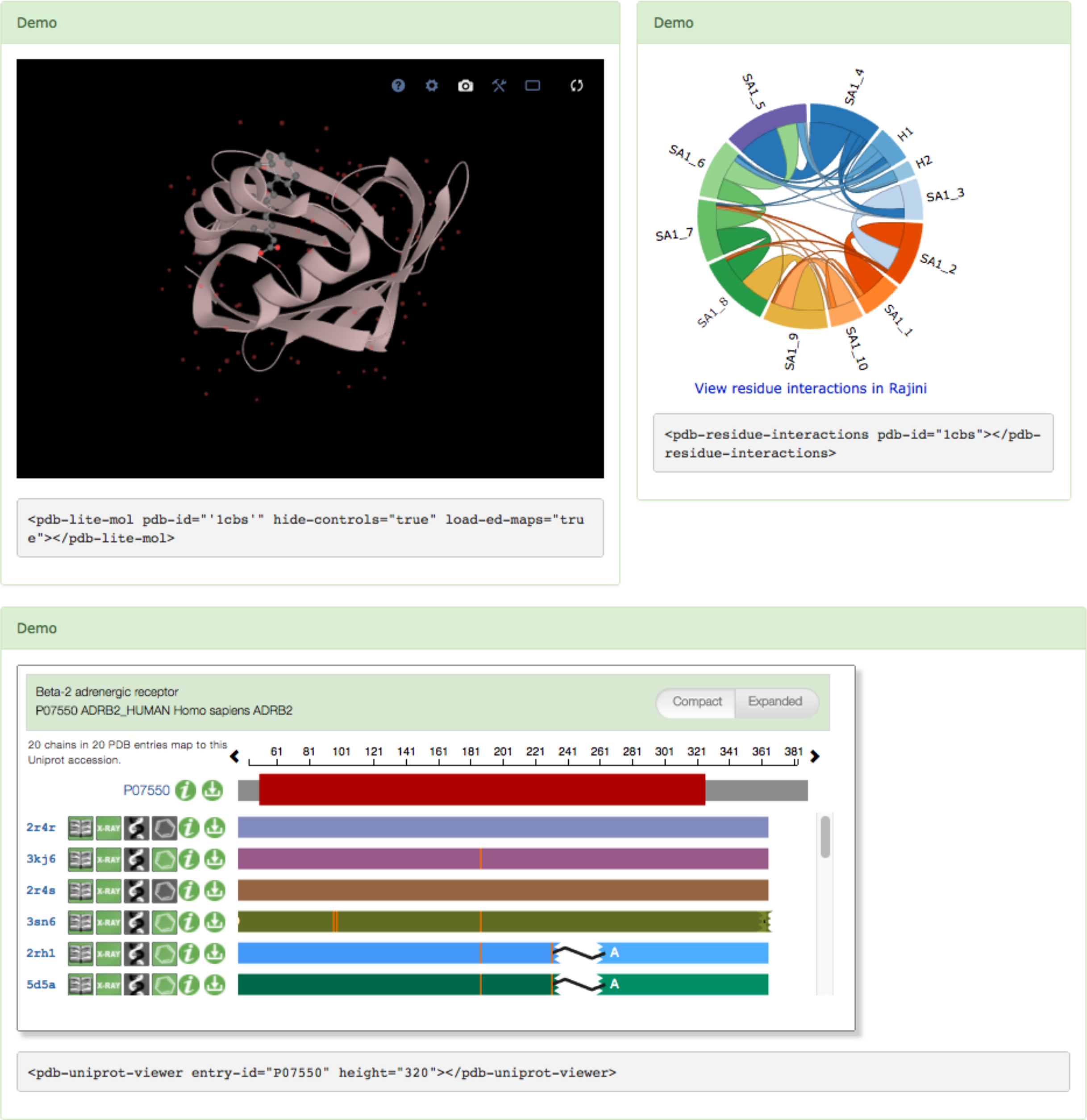
More...

Sequence family	
Haemocyan_bet_s	(6)
Staph_haemo	(1)

Figure 1.2.5.11: PDBe autocomplete feature showing various results for the query “haemo”

Upon selection of an entry, the web components are initialized with the corresponding PDB ID providing detailed information about the selected structure in an interactive manner through an easily understood user interface. The available web components include:

1. **PDB Links** containing links to all the wwPDB resources for a given PDB entry.
2. **PDB Prints,** a collection of PDB logos displayed in a specific order, where each icon represents a well-defined category of information.
3. The **PDB Topology Viewer**, which depicts the secondary structure of a protein in a 2D representation, taking into account the interactions of these secondary structure elements
4. The **Sequence feature View**, which shows a linear representation of the sequence of the protein and depicts value-added annotation from SIFTS. This includes residue-level mapping to UniProt, sequence families (Pfam), structure domains (SCOP, CATH), mutations, binding-site residues, structure quality and secondary structure.
5. The **PDB UniProt Viewer** (UniPDB), which displays all PDB entries mapped to a UniProt code, indicating their coverage and deviations from the UniProt sequence.
6. 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.
7. The **PDB Residue Interactions** component which is based on the residue contacts viewer in Rajini and the number of atomic contacts between each of the secondary structural elements (helices and sheets) in a protein.
8. The **PDB 3D Complex** component, which gives a compact summary of the results for a particular PDB code and assembly from the PDB 3D complex server.
9. **LiteMol** a powerful, streamlined structure viewer which enables a PDB structure to be explored within a browser rather than requiring pre-installed molecular graphics software.



```
<pdb-lite-mol pdb-id="'1cbs'" hide-controls="true" load-ed-maps="true"></pdb-lite-mol>
```

```
<pdb-residue-interactions pdb-id="1cbs"></pdb-residue-interactions>
```

Beta-2 adrenergic receptor
P07550 ADRB2_HUMAN Homo sapiens ADRB2

20 chains in 20 PDB entries map to this Uniprot accession.

```
<pdb-uniprot-viewer entry-id="P07550" height="320"></pdb-uniprot-viewer>
```

Figure 1.2.5.12: Demo of various web components including the LiteMol structure viewer (top left), the Residue Interactions Viewer (top right), and the UniProt Viewer (bottom)

Task 5.4 – Customized end-users VMs (STFC, all)

There were created 3 types of templates of virtual machine (VM) or container.

1) Virtual Folder VM - is based on CernVM, where operating system (currently Scientific Linux 7.3) is maintained and distributed continuously by CernVM project. VM templates of the Virtual Folder appliance are being released at appdb.egi.eu (current version 17.05) image list specific for the enmr.eu VO, so its deployment to EGI FedCloud requires user authentication with this VO. The VM boot time was reduced significantly by 60-70%. Binary updates of Virtual Folder are distributed continuously via /cvmfs/west-life.egi.eu. Deployment to EGI Federated Cloud is now available at INFN-PADOVA-STACK (IT), IFCA-LCG2 (ES) and CESNET-METACLOUD (CZ) sites. The template is registered in the appdb.egi.eu Intermittent bug of VM contextualization was reported to CernVM - workaround discovered and applied

https://sft.its.cern.ch/jira/projects/CVM/issues/CVM-1207. A technical plan to move /cvmfs/west-life.egi.eu under /cvmfs/wenmr.egi.eu (the default CVMFS path for the enmr.eu VO already configured in all sites supporting that VO) is being discussed and prepared.

2) INFN worked in collaboration with UU, CIRMMP, LIP, Indigo-DataCloud project and EGI-Engage "Accelerated Computing" task in order to develop a solution for running DisVis and PowerFit applications on docker containers which could be executed on grid worker nodes hosting GPGPU cards through a preview release of a CREAM-CE supporting accelerators. INFN also developed scripts to instantiate GPGPU-enabled VMs at IISAS (hosting NVIDIA™ Tesla K20m GPUs) and CESNET (hosting NVIDIA™ M2090 GPUs) sites of the EGI Federated Cloud, and to install on them the latest NVIDIA™ drivers and the DisVis and PowerFit software. The Partner CIRMMP recently installed an OpenStack private cloud on their premises. It is hosted on a blade with four AMD OPTERON processors with a total of 64 cores and 128 GB of RAM. It is connected to a iscsi SAN with 1.5 TB storage space.
3) ScipionCloud VM - A first version of the ScipionCloud virtual appliance (see M5.3) has been released. ScipionCloud represents the porting of the complete Scipion workflow environment for cryo EM image processing to private and public clouds (i.e., the European Federated Cloud and Amazon), including ways to speed up visualization for most popular cryo EM packages and GPU support.

1.2.5.4 Progress Towards Deliverables And Milestones

Milestones and Deliverables of WP5 have been met or are on schedule.

Number	Title	Lead	Type /planned delivery date	Progress
D5.1	Project portal	LUNA	Month 3	Delivered. The web representation of West-Life is publically available at https://west-life.eu
D5.2	Overview (baseline) of services and portals to be integrated into the new VRE (Month 4)	UU	Public report Month 4	Delivered. http://internal-wiki.west-life.eu/w/index.php?title=File:Overview_(baseline)_of_services_and_portals_to_be_integrated_into_the_new_VRE.docx
D5.3	Prototype of the new VRE portal functionality (Month 6)	Luna	Month 6	Delivered. http://internal-wiki.west-life.eu/w/index.php?title=File:Prototype_of_the_new_VRE_portal_functionality.docx
D5.4	Report on activities of the Helpdesk	Instruct	Public report Month 18	So far, a survey of the different means used to provide support for West-Life associated services has been compiled and numbers for the report are currently being gathered.
D5.5	DVRE-integrated PDBe search and query API's (Month 18)	EMBL	Public report Month 18	Delivered http://internal-wiki.west-life.eu/w/index.php?title=Deliverables

D5.6	Report on available VMs with associated documentation/use case for each of them	STFC	Public report Month 24	
D5.7	Report on access and usage statistics of the various services	UU	Public report Month 24	Update of statistics gathered for D5.2
D5.8	Report on the access and usage statistics	UU	Public report Month 36	Update of statistics gathered for D5.2 and D5.7
D5.9	Update report on activities of the helpdesk	Instruct	Public report M36	Update of report for D5.4

Milestone Number	Milestone Name	Estimated Date	Means of Verification
MS16	First version of new VRE portal online	4 / Delivered	The VRE portal implementation is available at http://portal.west-life.eu
MS17	First integrated grid- or cloud-enabled X-ray service	12 / Delivered	http://internal-wiki.west-life.eu/w/index.php?title=D6.1:CCP4_integration
MS18	First integrated, grid- or cloud-enabled cryo-EM service	18 / delivered	Towards M5.3 and in coordination with the MoBrain Competence Center we have released ScipionCloud as a virtual appliance on the EGI AppDB, available at https://appdb.egi.eu/store/vappliance/scipion.v1.0. The first version of the image provided Scipion 1.0.1 to be run on a single node plus remote desktop guacamole installed and configured to access the Virtual Machine through a Web Browser. A second version has been released on May 2017 that includes Scipion 1.1_beta release and GPGPU support. both can be downloaded and started on VirtualBox or used to instantiate a virtual machine on the EGI Federated Cloud. We worked together with MU towards Scipion Web Tools portal virtualization using the tools defined on the architecture to allow automatic deployment of the portal and Scipion software on EGI Federated Cloud. Furthermore and along this line we also collaborated with STFC to produce a version of Scipion that can be deployed using vagrant and is pre-installed on the Virtual Folder image.
MS19	First implementation of data catalogue in VRE	Month 18 / delivered	First implementation of data catalogue in VRE (Month 18) http://internal-wiki.west-life.eu/w/index.php?title=Data_catalogue

1.2.5.5 Deviations From Annex 1

There are no significant deviations from annex 1.

1.2.6 Work Package: 6 Data Management

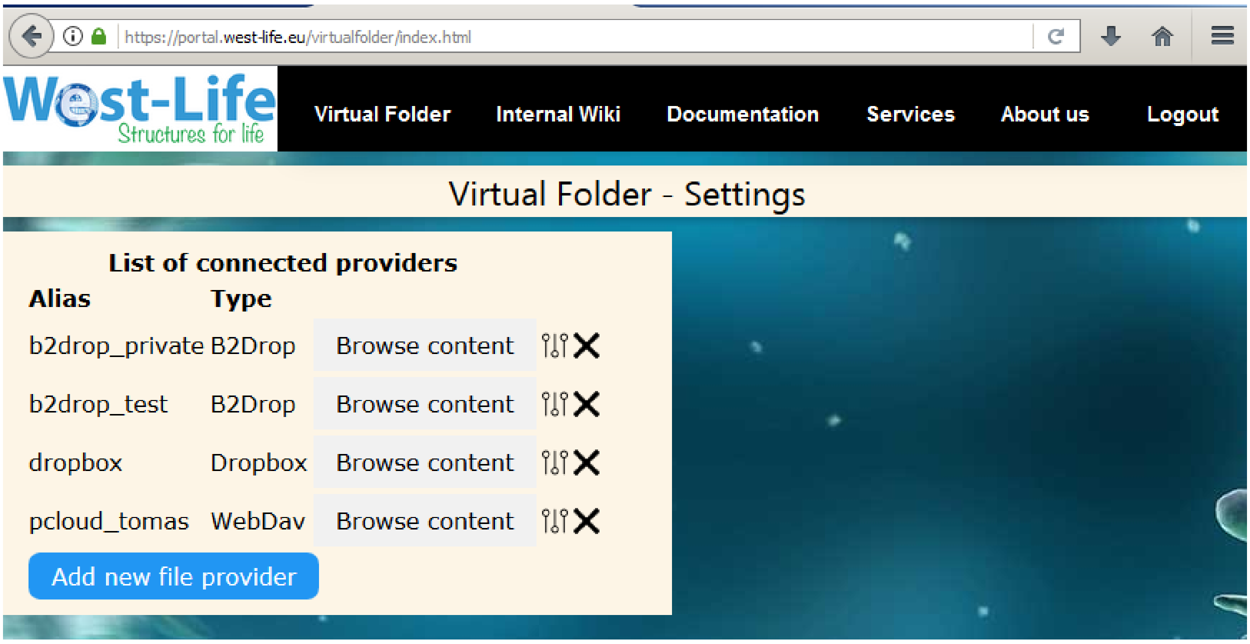
1.2.6.1 Summary and significant results

This WP is concerned with the development of a data management layer, to handle the diversity and distribution of research data. Much of the first reporting period has focussed on the creation of a Virtual Folder (VF) for unifying the different data sources available to the scientist. The VF is implemented as a Virtual Machine containing internal data, some data analysis software, and the means to access multiple external datasets. The VF exposes a number of services, allowing it to be deployed in various contexts. For example, it can be deployed by a user as a fully featured Virtual Machine, it can be accessed via the main West-Life portal after logging in to the site, or it can act as a file browser for structural biology services.

The VF has been integrated into the core architecture (WP4) and the VRE (WP5), as described above. Extensive documentation for users and developers is provided at https://h2020-westlife-eu.gitbooks.io/virtual-folder-docs/content/

The VF relies on access to multiple external data services. Accessibility of EUDAT services to structural biologists was improved by adding the ARIA identity provider used by Instruct to the B2ACCESS service.

A virtual machine was published in EGI's AppDB database, containing the software developed in this work package so far. This software was also integrated into the West-Life portal.



Virtual Folder - Settings

List of connected providers

Alias	Type
b2drop_private	B2Drop
b2drop_test	B2Drop
dropbox	Dropbox
pcloud_tomas	WebDav

Figure 1.2.6.1.1 Virtual Folder

1.2.6.2 Objectives and Tasks

This work package builds on existing infrastructure for storing and accessing data, to produce an application layer for data management suitable for the growing use of multiple techniques and multiple experimental facilities in structural biology research projects. As instruments improve and experimental methods diversify, structural projects are increasingly handling large numbers of datasets, which can be geographically distributed. This work package supports scientists in tracking, using, sharing, and discovering such datasets. This provides services to the portal which is under development in WP5.

Task 6.1

The effort builds on previous work by WeNMR and others in order to create a virtual folder view of scattered data (D6.1). In particular, it builds on the B2DROP and B2SAFE services provided by EUDAT, and on PaNData for photon and neutron facilities. This will be the first such effort to also address data management for the growing field of single particle electron microscopy.

CCP4 has begun development of a web portal for solving crystallographic structures. This will be integrated with the folder view and provenance view.

Use cases include: View files for project and keeping track of and process files on demountable storage devices. Standards to be used in this work include WebDAV, already supported by B2FIND, and iCAT-FUSE; Moonshot for authentication; and VoID, XRD, and EDAM for discoverability. The implementation will be compatible with existing CRIS repositories.

Task 6.2

For experimental facilities that are newly embarking on data management, we will provide a reference implementation of a repository that supplies suitable metadata to the portal (D6.2), matching the metadata standards to be devised in WP7. The use cases this will support are: registering a new project, and adding files to a project. The implementation will use the CERIF standard. It will be compatible with existing CRIS repositories, and also be capable of assigning an URI to a project if it is not yet recorded in a CRIS repository.

Task 6.3

We will then extend that to a provenance view showing the contribution of each sample, experiment, and dataset to the conclusions of the project (D6.5), using the PROV-O standard, including reporting PROV-O metadata from iCAT.

Task 6.4

We will facilitate collaboration between the computational modeling and structural biology community by providing the necessary data management, dissemination and analysis tools (D6.3) for predicted models. We use metadata standards to be developed in WP7 for description of predicted models, and implement the necessary data dissemination mechanisms (D6.4).

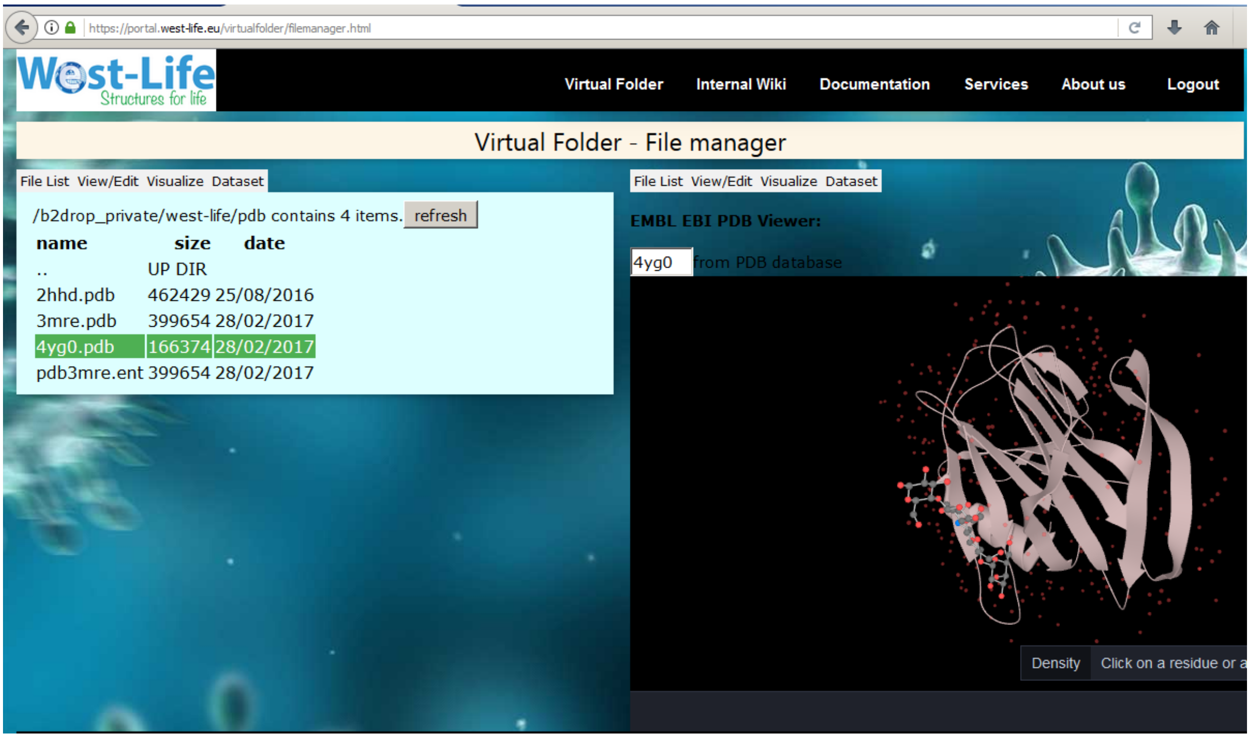
1.2.6.3 Work carried out in P1

D6.1 provides a consistent view of the files for a research project, regardless of the experimental facility in which they were obtained and regardless of their current location. This view is available through a web interface, and also as a mounted file system for access by programs which the user runs.

This work used the B2DROP service provided by EUDAT. In order to facilitate use of EUDAT services by structural biologists, we worked with ARIA and EUDAT to ensure that Instruct userids are accepted by EUDAT's authentication service B2ACCESS [https://www.structuralbiology.eu/update/news/publish-data-with-b2share/]. Additionally the integration was done with DROPBOX service and any other service providing access to files via WEBDAV standard API.

The user interface was designed and implemented in order to provide “File manager” view of scattered data, visualization of PDB data files with integration of LiteMol visualization tool and dataset web components originaly delivered in WP5. An integration with other third party portal was prepared using Javascript cross-origin communication feature and is demonstrated in File Picker component Fig. 1.2.5.2

and being implemented in DISVIS portal (WP5).

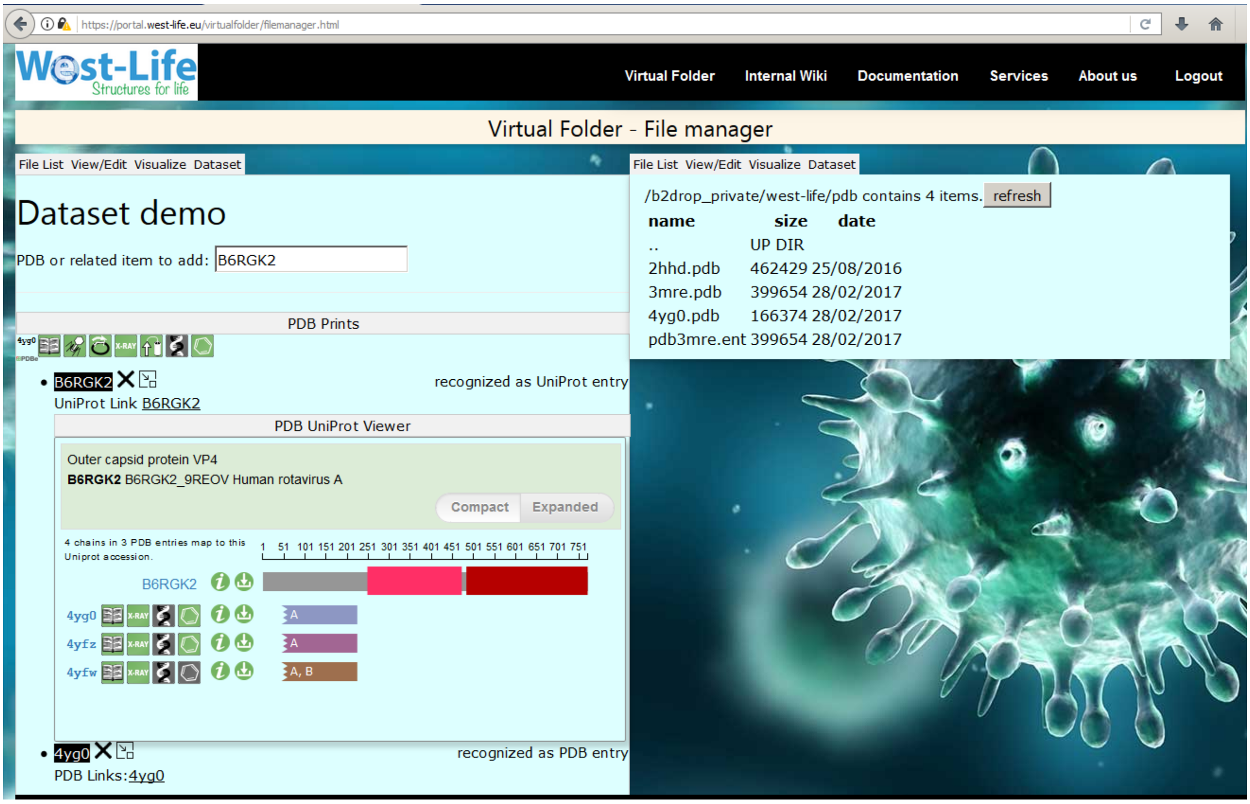


/b2drop_private/west-life/pdb contains 4 items.

name	size	date
..	UP DIR	
2hhd.pdb	462429	25/08/2016
3mre.pdb	399654	28/02/2017
4yg0.pdb	166374	28/02/2017
pdb3mre.ent	399654	28/02/2017

4yg0 from PDB database

Figure 1.2.6.1.2 File Viewer



Dataset demo

PDB or related item to add: B6RGK2

- B6RGK2 recognized as UniProt entry
  UniProt Link B6RGK2

  Outer capsid protein VP4
  B6RGK2 B6RGK2_9REOV Human rotavirus A

  4 chains in 3 PDB entries map to this Uniprot accession.

  B6RGK2
  4yg0
  4yfz
  4yfw
- 4yg0 recognized as PDB entry
  PDB Links:4yg0

/b2drop_private/west-life/pdb contains 4 items.

name	size	date
..	UP DIR	
2hhd.pdb	462429	25/08/2016
3mre.pdb	399654	28/02/2017
4yg0.pdb	166374	28/02/2017
pdb3mre.ent	399654	28/02/2017

Figure 1.2.6.1.3 PDBe API in use

The backend service was implemented in C# leveraging the Servicestack.NET framework to deliver REST

API of all stored configuration and metadata. Part of backend service is configured as a virtual directories of Apache web server delivering WEBDAV standard interface with reverse proxy mechanism to appropriate endpoint.

Next to the portal deployment, the virtual machine template was prepared with custom software and tools and mounted storages of virtual folder are accessible on file system level to any legacy software. Thus tools such as CCP4 suite or SCIPION suite can be launched via remote desktop and can access the files of virtual folder via local file system.

1.2.6.4 Progress towards Deliverables and Milestones

Number	Title	Lead	Type /planned delivery date	Progress
D6.1	Virtual Folder	STFC	Month 12	Delivered http://internal-wiki.west-life.eu/w/index.php?title=File:West-Life_D6_1.docx
D6.2	Repositories	STFC	Public report Month 24	Reporting Period 2
D6.3	Assembly queries	EMBL	Month 32	Reporting Period 2
D6.4	Provenance	STFC	Public report Month 133	Reporting Period 2

Milestone Number	Milestone Name	Estimated Date	Means of Verification
MS20	Virtual folder	20 / Already Delivered	Accepted by Luna as basis for VRE
MS21	Repository	24 / Reporting Period 2	Installed by some facilities. CERM has agreed to be the reference implementation.
MS22	*Metadata standards for complexes*	12/ dleivered	File:Report on existing standards.docx made available on project website
MS23	Query interface for complexes	Month 24 / Reporting Period 2	Report of user requirements and technical report on implementation plan

1.2.6.5 Deviations from Annex 1

There were some delays in hiring, therefore it was not possible to link iCAT to the Virtual Folder in this period. This work will be performed in period 2.

1.2.7 Work package: 7 Joint Research

1.2.7.1 Summary and significant results

The goal of this WP is to explore ways to extend, combine and enhance existing services, so that they can reach a wider audience. Others topics to be addressed were focused research on exploring Big Data approaches together with a number of standardization issues.

During this first reported period all Milestones and Deliverables have been reached, at the same time that work towards most other project Deliverables is proceeding within schedule (and even ahead of schedule).

1.2.7.2 Objectives and Tasks

Work Package 7, as a Joint Research Activity, is aimed at exploring new ways to use existing or close to existing services so that broader user communities will be reached. This overarching goal will be accomplished through 4 Objectives which have a one-to-one relationship with the 4 Tasks that are outlined in the following:

Task 7.1 – Extending and benchmarking existing web services (EMBL-HA, CSIC, STFC, NKI, MU, CIRMMP, Instruct, UU)

Four “exemplary” and very popular services were chosen to concentrate our efforts on: ProteinCCD, PDB-REDO, REFMAC and HADDOCK.

ProteinCCD is aimed at helping the user in the design of suitable genomic constructs to obtain good samples to be analyzed by the different structural biology techniques. Our goal in WP7 is to extend ProteinCCD analysis options as well as increase the range of information to be taken into account. This latter development requires system architecture changes to enhance performance and reliability.

PDB-REDO, in turn, automatically optimizes the refinement of atomic models from X-ray crystallography, tying together different software engines with newly developed tools and decision making algorithms. In WP7 we aim at increasing its performance and, at the same time, substantially enlarge the user base thanks to integrating the database of ready-made alternative, and in most cases improved, versions of models already deposited in structural databases (PDB), with the web server that allows crystallographers to optimise the structure model on which they are currently working.

REFMAC, which is a very widely used core model refinement service, accessed on its own through CCP4 and through model building services with ARP/wARP. ARP/wARP is one of the most widely used systems in MX, providing an automated model building of protein models as well as their complexes with DNA/RNA and bound ligands. In this case, the goal is to further enlarge its user base, so that model building and refinement at an atomic level does not only address X-ray diffraction data, but also cryo EM and NMR data.

Finally, HADDOCK is one of the best known services for biomolecular docking (for High Ambiguity Driven biomolecular DOCKing). HADDOCK takes as input atomic models as well as spatial constraints, and the challenge for WP7 is to extends its computational approach and the service to directly take into account cryo EM maps (i.e., 3D density maps, without atomic modeling).

Task 7.2 – Combining existing services into new workflows (CSIC, EMBL-EBI, STFC, NKI, MU, CIRMMP, UU)

In this Task the focus is on quality measures, selecting two fields of activity. The first one is cryo EM, as the case of a very rapidly expanding technique that requires to define and integrate a new pipeline geared towards quality measures. The second one is aimed at the quality analysis of predicted complexes.

As for cryo EM, local resolution and other approaches were considered, which implied the use of tools such as ResMap and others, some included in Scipion Web Tools. As for the evaluation of predicted complexes, the goal was to assist the CAPRI community and to contribute to new methods development.

Task 7.3 – Handling and Mining Big Data (STFC, (CSIC, MU))

This Task was strongly motivated by the recent collaborative involvement of IBM Research into Structural Biology, which was supported by a strong letter of support to the project. The goal was to explore a number of Big Data technologies for handling and interpreting data in West-Life, using where appropriate IBM know-how and tools.

Originally, this Task was very open, and it is being further defined along the project, as indicated in the next section of this progress report. Ideas in the original proposal covered the use of MapReduce formulations, the use of flash storage, applications for streaming data, and IBM Watson for cognitive computing. In addition to the latter, we are also considering more general applications of machine learning and natural language processing.

Milestone 7.7 is due in month 24, and Deliverable 7.8 in month 30. Thus, this task is scheduled for later in the project, and progress to-date has mainly concerned discussions of ideas. The aim is to prototype innovative approaches which might be quite different to current practice. These prototypes will be disseminated to project partners, but will not necessarily be included in production services during the lifetime of this grant.

Task 7.4 –Assessing and Extending metadata formats (STFC, EMBL-EBI, NKI, MU, CSIC, CIRMMP, Instruct, UU)

The main activity here was the survey of available metadata standards relevant to the set of services offered by West-Life, identifying existing standards that could be used in the services as well as “gaps” to be addressed. The project services are to be updated using these standards.

Milestones 7.8 and 7.9 are due in months 24 and 30 respectively, and Deliverable 7.9 is due in month 30. Again, most work on this task will be in the second period of the project.

1.2.7.3 Work carried out in P1

Task 7.1 – Extending and benchmarking existing web services (EMBL-HA, CSIC, STFC, NKI, MU, CIRMMP, UU)

ProteinCCD. The original ProteinCCD application was a Java applet and a popular tool for many scientists, with 250-350 users per month. Java applets, however, have security issues and have become deprecated. We therefore re-implemented ProteinCCD2 as a web application with a Flask/Biopython backend and a Bootstrap/Javascript front-end. The new implementation we delivered eliminated security concerns, making the application available to any device able to run a web browser.

The new ProteinCCD2 included numerous new features that were reported in Deliverable 7.2; briefly: a) enabled parallel processing of server requests; b) a more efficient interface for construct design was made available; c) the internal database is checked on the fly for vectors compatibility, increasing user awareness of the experimental possibilities; d) better support for restriction based cloning is provided; e) local execution of algorithms improves response time; f) new servers for meta-analysis have been added (e.g. HNN , MLR, DPM, PREDATOR, IUPRED, GLOBPLOT and COILS) for easy saving of predictions, primers and resulting peptides has been implemented to enable efficient book keeping. Importantly, we also made excellent progress towards final deliverables, as reported in Milestone 7.1, making progress towards ranking design in ProteinCCD. We ultimately want to aid the users in making optimal decisions and for that we aim to add solubility and crystallizability scoring and ranking algorithms from available servers. We have split the analysis and ranking in three tiers: a) use algorithms from the BioPython library to calculate the molecular weight, isoelectric point and absorption coefficients for all designed constructs; b) for providing solubility information we evaluate from external servers; c) for providing crystallizability information we also chose external servers. The test version of these implementations are already underway and we are confidently on schedule for timely delivery of Deliverable 7.3.

PDB-REDO: The computational backend software of PDB-REDO web server was parallelised to reduce the running time of user jobs and thus improve the user experience (Deliverable 7.1, delivered). The web server engine was rewritten to incorporate ARIA based single sign-on and an API was created to allow other web servers in West-life to submit jobs in a multi-server workflow (Task 7.2). The web server and the PDB-REDO databank were moved to a single domain (pdb-redo.eu) to allow the development of new modes of use that rely on the integration of the databank and the web server (milestone 7.2, delivered). The user interface of the databank was redesigned to provide a more clear overview of the available data and to allow user to better judge the effect of model optimisation by PDB-REDO.

REFMAC: Remote access to the REFMAC engine will be provided within the next-generation Web Service of the automated model-building suite ARP/wARP. The workflow and i/o of the Refmac web service has been defined (Milestone 7.3, accomplished). In brief, at first, the input data is validated for consistency and any issues are fixed if appropriate and reported to the user. In case of EM data, next, the map and the coordinates are converted to a minimum pseudo-crystallographic asymmetric unit-cell that in some cases may substantially reduce computation time. The map in a new coordinate frame is then used to calculate structure factors in the MTZ format, which is required by REFMAC. Finally, a REFMAC-based reciprocal-space refinement or a combined iterative ARP/wARP-REFMAC model building is performed depending on the users choice and the characteristics of the data. At the last step, a refined model is back-transformed to the original reference frame and provided to the user in PDB and mmCIF file formats.

HADDOCK: HADDOCK heavily relies on two distinct frameworks to work. The core one is its main and standalone framework that handles a user job from the start to the end, making use of HTC resources for the core calculation. The second one consists in the web server that communicates back-and-forth with the standalone framework. As reported for Milestone 7.4, the different components and interaction between the two frameworks is illustrated in the Figure below:

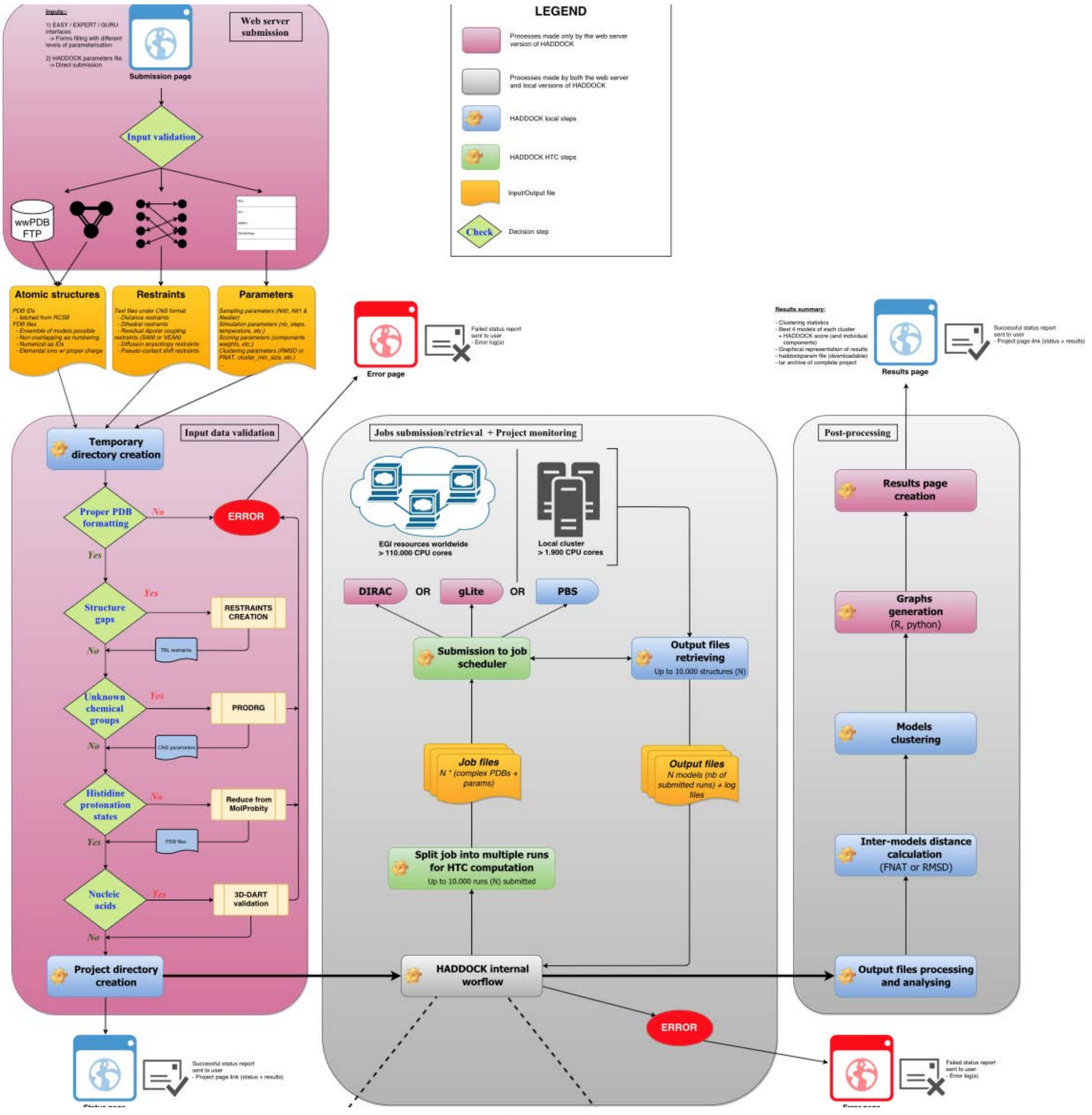


Figure 1.2.7.1.1_Workflow of the HADDOCK web server.

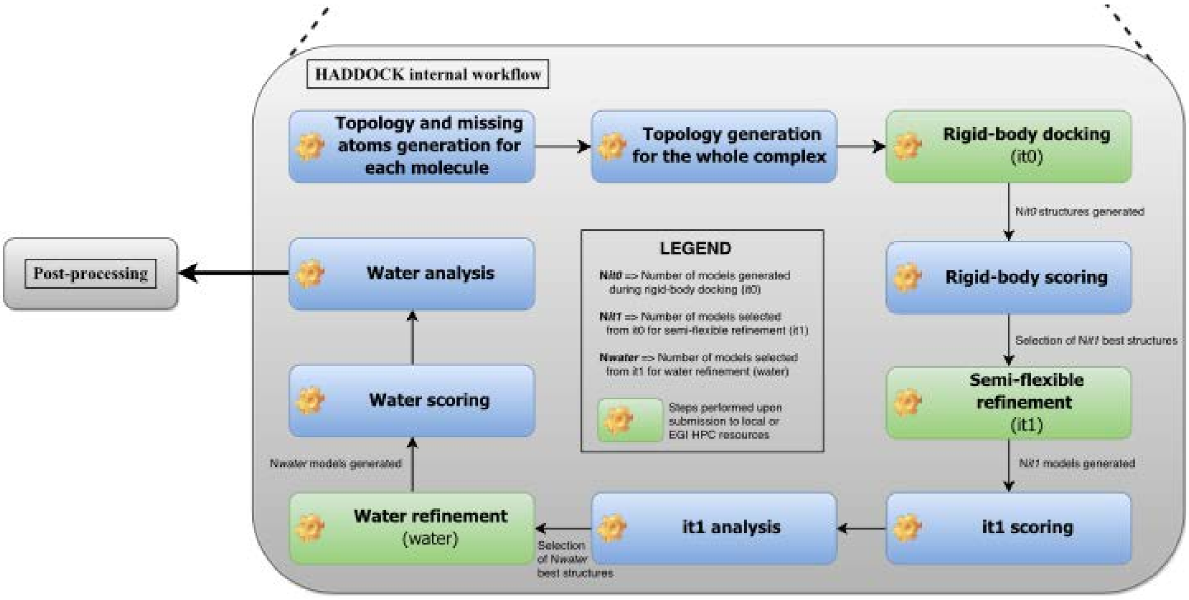


Figure 1.2.7.1.2_HADDOCK internal workflow (corresponding to a local installation of HADDOCK)

We started the development of the new version of HADDOCK, HADDOCK3.0, which will bundle a list of new features both at the CNS core (simulation engine) and front-end levels and support cryo-EM restraints. These developments involve the standalone version and the web server as well. Some preliminary steps have already been made towards a new server workflow to handle HADDOCK pipeline that would get rid of a former local custom library (Spyder), developed by a former lab member. This Spyder layer is currently performing the first validation in the server workflow before calling the Python layer.

Together with this major update, we also started the complete rewriting of the web portal and server. The web server will now rely on the Flask framework (http://flask.pocoo.org/), successfully used for two recent web portals, DisVis and PowerFit. Flask will allow for a tighter integration of the standalone workflow within the web server one and will bring new capabilities at the web client level. Among them, we target more interactivity, a better visualization of data, a faster validation, possibility to login to get a personalized experience, possibility to monitor current jobs, etc. It will also allow a much faster inclusion of new features in the web portal, something which requires a major effort in the current Spyder setup. This will be a major achievement in HADDOCK development and will be broadly advertised through our usual channels of communication. A complete documentation is also planned to be released together with this new version.

Besides extending the services mentioned above, work has also been carried out on cryo EM services, specifically on Scipion (Scipion Web Tools) and XMIPP packages. The work has been conducted mostly by MU, in coordination with CSIC.

In this way, performance-critical sections of XMIPP code were identified, focusing on one of the most obvious hotspot, which is the 2D image alignment procedure. After thorough analysis, the code was refactored to benefit from the use of vectorizing compilers. These changes produced more than 2x speedup of the code running on conventional CPU. The new version was also thoroughly tested in terms of equivalence with the previous version. It will be released to public with the next version of XMIPP. We note that refactoring is the first necessary step to achieve high performance on hardware accelerators. The new code will also be evaluated on Xeon Phi, where immediate further speedup is expected. Further, we will consider porting the code also to CUDA or OpenCL in order to leverage GPU acceleration.

Task 7.2 – Combining existing services into new workflows (CSIC, EMBL-EBI, STFC, NKI, MU, CIRMMP, UU)

Combining services in cryo-EM:

Several workflows for quality assessment of cryo-EM data have been developed during this first reporting period. Local map resolution is certainly an important quality parameter for cryo-EM maps, which has been addressed in two different ways: (1) by modifying the popular program ResMap to become part of Scipion Web Tools (SWT) and, consequently, West-Life, as initially planned, and (b) by incorporating other new approaches developed since this then, such as MonoRes. Additionally, we have explored another processing step (not initially considered in the project) which is at the core of the cryo-EM ability to obtain 3D density maps of biological macromolecules, as it is the finding of the relative orientations of the hundred of thousands of individual macromolecules that are being imaged and that contribute to the final 3D map. Indeed, this step it is one of the most important sources of error in the reconstruction process. As a result of this latter work, a new workflow combining the result of two alignment reliability methods has been successfully incorporated to West-Life repertoire. The figure below shows the graphical display in Scipion Web Tools of the new workflows.

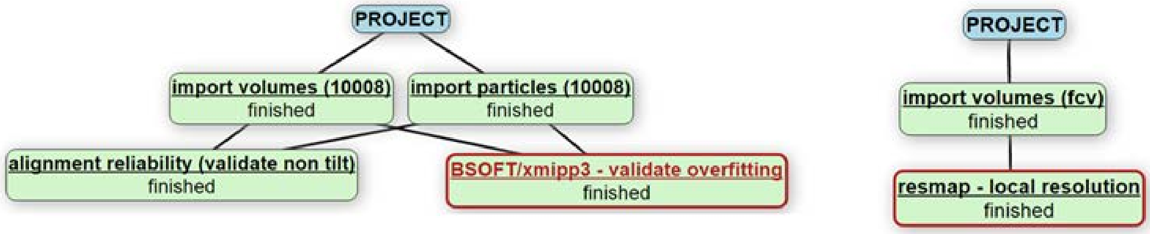


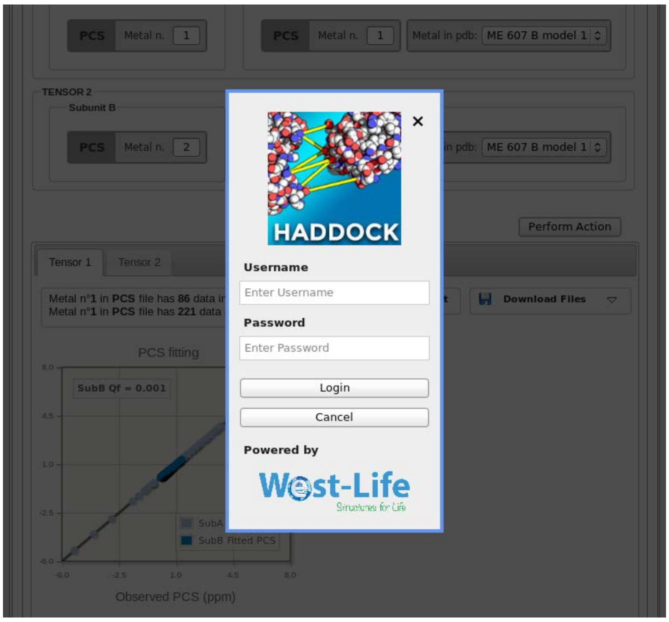
Figure 1.2.6.2.1

In this context, UU has extended its PowerFit service to accept density maps from SWT, enabling in this way the possibility of connecting the 2 services. SWT have adapted its volume viewer to offer the users the option to send the density map to PowerFit.

Regarding cloud-based services, SWT and the whole of Scipion framework have been ported to the European Federation Cloud as well as to public clouds (Amazon). In turn, MU has implemented a prototype of automatic deployment of the SWT on the European Federated Cloud using cloud orchestration and software configuration tools (Cloudify and Puppet) and successfully configured the validation tools to be run in a cluster. Based on the positive experience with automatic cloud deployment of SWT, combined cloud deployment of SWT and PowerFit was designed. In this setup, both tools will be merged to co-exist in a single virtual cluster, allowing to interchange data smoothly, without the need of additional inter-service communication.

Refining protein complexes obtained from FANTEN with HADDOCK.

We are currently testing the setup of a bridge between FANTEN (http://abs.cerm.unifi.it:8080/), a rigid-body refinement protocol driven by Pseudocontact shifts and Residual Dipolar Couplings measurement from NMR (Rinaldelli M, Carlon A, Ravera E, Parigi G, Luchinat C. J Biomol NMR., 61, 21-34 (2015). FANTEN: a new web-based interface for the analysis of magnetic anisotropy-induced NMR data.), and the flexible refinement protocol of HADDOCK, UU docking software. First tests showed that a straightforward pipeline could be made from FANTEN output models to HADDOCK input requirements in order to generate high quality models of protein-protein complexes. We are currently testing the performance of the whole protocol on specific test cases. In parallel, discussions are in progress to decide on the technical details of the implementation.



Integration of pseudocontact shifts and residual dipolar couplings in REFMAC

CIRMMP is active in the integration of NMR-based restraints, particularly restraints derived from so-called paramagnetic NMR data, in the REFMAC software, in collaboration with the team of Dr. Garib Murshudov. The corresponding software routines constitute an extension that we called REFMAC-NMR (doi: 10.1107/S1399004713034160). The developments achieved since the beginning of West-Life regard:

- the use of non-crystallographic symmetry constraints during the estimation of the paramagnetic susceptibility (for pseudocontact shifts) and/or of the orientation (for residual dipolar couplings) tensor

- the use of orientation and anisotropy restraints for multi-domain proteins as well as macromolecular complexes

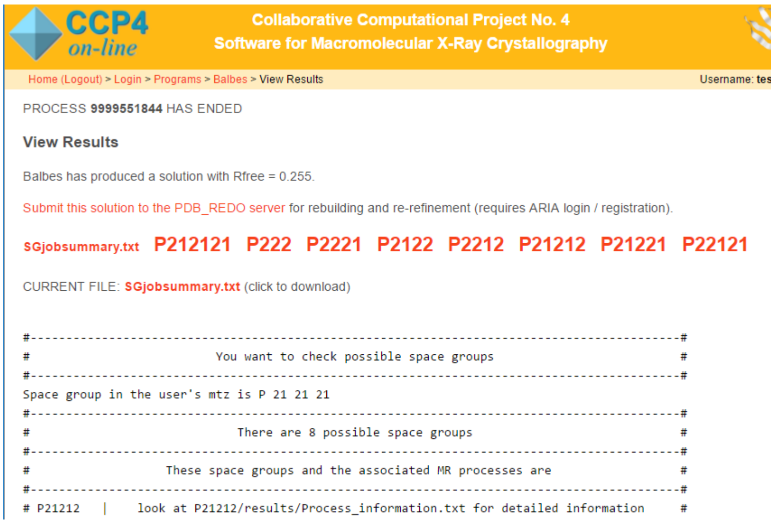
- the analysis of residual dipolar couplings for chemical bonds in aminoacidic side chains containing methyl groups (isoleucine, leucine, valine). These side chains can take different rotameric states (also visible in the crystal). Thus, when alternative conformations are used to fit the electron density, it is possible to carry out a joint refinement with the residual dipolar couplings, which are interpreted using an "ensemble" approach. That is, the back-calculated residual dipolar couplings are computed as the weighted average (through the occupancy factor of the altConf) of the coupling obtained for each alternative rotamer.

Connecting crystallographic structure solution pipelines with model re-building, refinement and validation services

CCP4Online has several pipelines for solving crystal structures via Molecular Replacement, including Balbes, MoRDa and MrBUMP. The ARP/wARP service can re-build and extend an initial solution, while the PDB_REDO service can re-build and refine a structure. Previously, it was possible to request that the output of Balbes be sent to the ARP/wARP service, but otherwise the responsibility for coordinating jobs on these services lay with the user.

We have implemented changes to the CCP4Online pages (currently on the development server) allowing the user to select one or more possible solutions, and send these directly to ARP/wARP or PDB_REDO or both. Similarly, users of ARP/wARP can select results to be sent to the PDB_REDO service. Communication between the services is handled by newly-developed APIs. The user can make multiple submissions, with those coming from the same structure solution job linked by a group ID. The ARP/wARP and PDB_REDO services authenticate users via ARIA, and a login prompt is presented if the user is not already signed in, in the current browser session.

Future work will increase the level of integration. In particular, the Virtual Folder (WP6) will be used to manage input and output files, and be made visible to all 3 services.



PROCESS **9999551844** HAS ENDED

View Results

Balbes has produced a solution with Rfree = 0.255.

Submit this solution to the PDB_REDO server for rebuilding and re-refinement (requires ARIA login / registration).

SGjobsummary.txt **P212121** **P222** **P2221** **P2122** **P2212** **P21212** **P21221** **P22121**

CURRENT FILE: **SGjobsummary.txt** (click to download)

```
#-------------------------------------------------------------------------------------------#
#                          You want to check possible space groups                          #
#-------------------------------------------------------------------------------------------#
Space group in the user's mtz is P 21 21 21
#-------------------------------------------------------------------------------------------#
#                             There are 8 possible space groups                             #
#-------------------------------------------------------------------------------------------#
#                    These space groups and the associated MR processes are                 #
#-------------------------------------------------------------------------------------------#
# P21212   |    look at P21212/results/Process_information.txt for detailed information    #
```

Task 7.3 – Handling and Mining Big Data (STFC, (CSIC, MU))

As mentioned above, this task is scheduled for later in the project, and progress to-date has mainly concerned discussions of possibilities. Specific ideas are being collated on the West-Life wiki, under the WP7 pages, and we are discussing possible prototype projects with IBM Research who have staff located at STFC Daresbury Laboratory.

We have begun to scope a project on the use of natural language processing for interpreting journal articles associated with PDB and EMDB entries. EMBL-EBI has a long-standing need to extract more information from such publications, for example, which other relevant proteins/small molecules are mentioned/described in the publication. In addition to textual information, we would like to understand which figures are relevant, for example, from an analysis of the figure legends. We would like to expand this analysis out to other publications which may give supporting information. These additional publications may be identified through citations, common authors or through referencing the PDB/EMDB ID.

Given a set of relevant publications, we would train our NLP model to identify particular technical terms, and aim to deliver an intelligent search engine. This means (for example) that you could ask what experiments have been done on the sample (beyond the obvious xtal/EM) and the model will have some internal idea what words would relate to an experiment. There would be significant overlap with other NLP projects in biology, e.g. in identifying protein / molecule names in text, and the NLP model is expected to have wide applicability.

Other Big Data approaches are being discussed (see the West-Life wiki) but are not yet as well defined.

Task 7.4 –Assessing and Extending metadata formats (STFC, EMBL-EBI, NKI, MU, CSIC, CIRMMP, Instruct, UU)

This task is related to Milestone M22 of WP6 “Metadata standards for complexes” which was delivered in month 12. That report looked at existing standards for describing macromolecular complexes, which are central to West-Life in that they are typically characterised by multiple experiment types. Task 7.4 will be build on this work, looking at a wider range of topics, and making recommendations for new or extended standards.

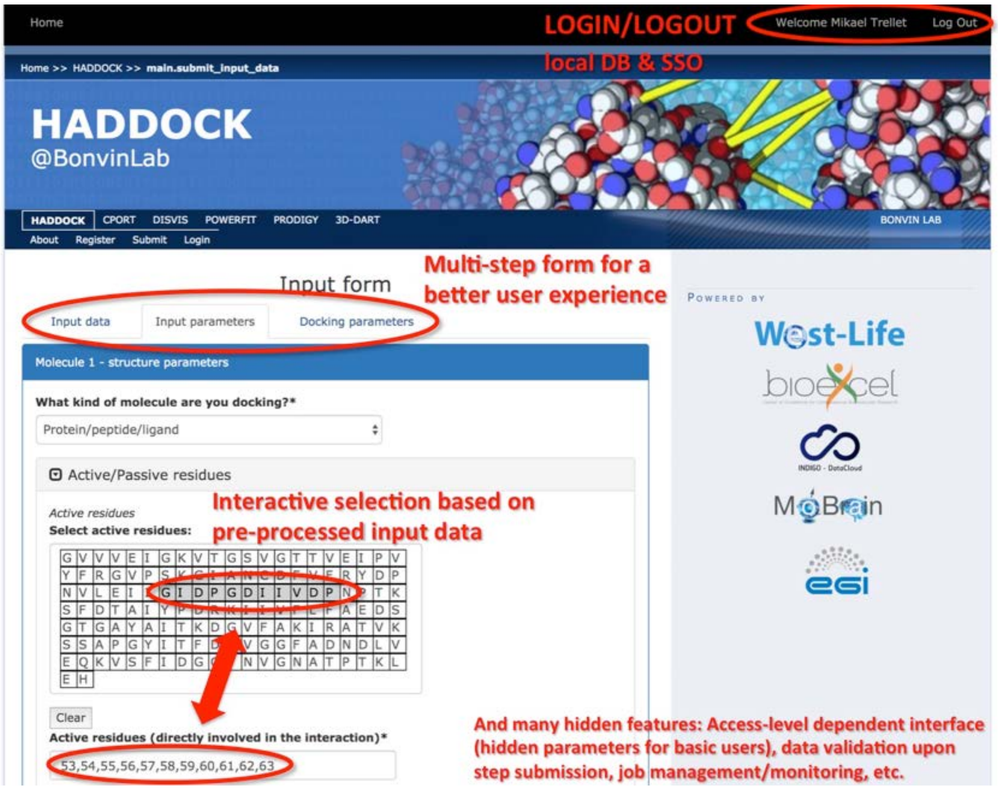
Initial work is being collated on the West-Life wiki, under the WP7 pages.

1.2.7.4 Progress towards Deliverables and Milestones

Number	**Title**	**Lead**	**Type /planned delivery date**	**Progress**
D7.1	Multi-core implementation of PDB REDO server	NKI	Public report Month 6	Delivered http://internal-wiki.west-life.eu/w/index.php?title=File:Multi-core_implementation_of_PDB_REDO_server.docx
D7.2	ProteinCCD with new analysis options	NKI	Public report Month 9	Delivered http://internal-wiki.west-life.eu/w/index.php?title=File:West-Life_D_7_2.docx
D7.3	ProteinCCD with construct scoring and ranking	NKI	Pubic report Month 24	Reporting Period 2. Excellent progress towards this Deliverable is being made, with an initial version (still non feature-complete) already available on-line.
D7.4	A REFMAC web service for EM and NMR	EMBL Month 24	Public report Month	Reporting Period 2 Work at EMBL-HA is proceeding well, specially in its extension to EM. Work on REFMAC-NMR is currently being analyzed.

D7.5	A HADDOCK server for EM	UU	Public Report Month 24	Reporting Period 2 See below.
D7.6	EM quality assurance workflow	CSIC	Public Report Month 24	Reporting Period 2 Progress towards this Deliverable is ahead of schedule, with two new cryo-EM quality assurance workflows already available, as indicated in detail in the previous section.
D7.7	Quality analysis workflow for predicted complexes	EMBL	Public report Month 30	Reporting Period 2 We plan to start work soon after April 2017.
D7.8	Report on prototypes constructed using Big Data approaches.	STFC	Public report Month 30	Reporting Period 2 As indicated in the previous section, the precise scope of the prototype is currently under evaluation/discussion in West-Life wiki.
D7.9	Report on existing metadata standards, and proposals for new vocabularies	STFC	Public report Month 30	Reporting Period 2 Compilation of information is proceeding through the project wiki pages.

The web interface of HADDOCK is currently undergoing a complete rewrite in Flask. This involves porting all current features of HADDOCK2.2 web server in Flask together with the integration of new ones, in particular the support for cryo-EM restraints and coarse grain models, which will be introduced with HADDOCK3.0. Due to the complexity of the web server, which currently provides 7 interfaces to run a docking simulation with different level of control over the parameters and different input data, we are organizing the development of the web server in a bottom-up manner where independent and simple tasks are progressively implemented, linked and extended. An overview of the first developments made for HADDOCK3.0 on Flask are illustrated in Figure 1.2.7.1.



LOGIN/LOGOUT

local DB & SSO

Multi-step form for a better user experience

What kind of molecule are you docking?*: Protein/peptide/ligand

Interactive selection based on pre-processed input data

Active residues (directly involved in the interaction)*: 53,54,55,56,57,58,59,60,61,62,63

And many hidden features: Access-level dependent interface (hidden parameters for basic users), data validation upon step submission, job management/monitoring, etc.

Figure 1.2.7.4.1 - New HADDOCK3.0 submission form interface

Considering that this complete redesign of the web portal represents a significant amount of work, there may be some delay in this Deliverable.

Milestone Number	**Milestone Name**	**Estimated Date**	**Means of Verification**
MS24	Ranking in ProteinCCD	Month 12 Delivered	File:Specifications for new ranking services in ProteinCCD.docx
MS21	Integration of PDB_REDO server and database	Month 18 Delivered	The PDB-REDO server and databank were moved together to new domain (pdb-redo.eu). A new web server engine was written to allow more flexible presentation of structural (meta)data. Users can now use their ARIA identity to login and to run server jobs. Additional functionality, such as user-specific presentation of databank (meta)data or the possibility for signed-on users to initiate updating of existing databank entries, are being implemented.
MS26	The workflow and i/o of the Refmac web service defined	Month 12/ delivered	The workflow and associated parameters and dat [File:M7.3.docx]()
MS27	The workflow and i/o of the HADDOCK web service defined	Month 24 / Reproting Period 2	The workflow and associated parameters and data model published [File:Milestone 7 4.pdf]()
M28	The EM Quality Assurance workflow defined	Month 12 Delivered	The workflow and associated parameters and data model published on the consortium web site. [File:M7 5.pdf]()
M29	The quality analysis workflow for predicted complexes defined	Month 24 Reporting Period 2	There is a well defined method for analysing the computationally predicted complexes that is used by the CAPRI ([pdbe.org/capri]()) community. We have contacted the developer and will get the necessary code which we plan to implement at PDBe as a web service. This will include different ways to analyse protein-protein and protein-peptide complexes that are predicted using computational methods and possibly with additional experimental data.

M30	Big Data software introduced	Month 24 Reporting Period 2	We are holding discussions over summer 2017 to discuss prototype Big Data projects. We are talking to IBM Research staff located at STFC Daresbury Laboratory, and investigating the use of IBM Big Data Software. We are also looking into relevant open source software. We are on course to make recommendations for this Milestone in month 24.
M31	*Existing metadata services identified*	Month 12 Delivered	Existing standards are currently being collated on the West-Life wiki. There will be a meeting of the wwPDB PDBx/mmCIF Working Group on Monday 17th & Tuesday 18th July at EMBL-EBI, which is expected to provide relevant input to this milestone.
M32	Existing services updated to existing data types	Month 30 Reporting Period 2	In collating standards for M7.8, we are also considering which West-Life services these are appropriate to. These links will also be recorded on the West-Life wiki, and then communicated to service providers.

1.2.7.5 Deviations from Annex 1

Deliverable 7.5 may be late, for reasons explained above.

1.3 Impact

This section was completed online.

1.4. Access provisions to Research Infrastructures

Not applicable to this project.

1.5 Resources used to provide access to Research Infrastructures

Not applicable to this project.

2. Update of the plan for exploitation and dissemination of result

No update is needed.

3. Update of the data management plan

No update is needed. Note that this project does not generate experimental data.

4. Follow-up of recommendations and comments from previous review(s)

The reviewers of the grant application made three points that are addressed here:

- A plan and a report on sustainability are envisioned, however, no related KPIs are provided for the effectiveness of IPR management.
- The VRE will benefit from being anchored in major research institutions.
- Activities regarding sustainability are addressed in the risk plan but they could have been given more emphasis.

IPR Management

The intellectual property generated by West-Life is in the form of computer code, both to operate web sites and other infrastructure, and as extensions to existing algorithmic code.

The newly developed codes are open source. The extensions to algorithmic codes which are being made for combined techniques are subject to the existing licences for these codes. In some cases these too are open source. In other cases they use an IP management model that has proved effective in structural biology: they are free for academic use, and revenue from industrial use is used to help sustain the software.

We will track the use of these two different models within the partnership.

The VRE will benefit from being anchored in major research institutions

The structural biology web services which add up to make the VRE are hosted by major research institutions including CSIC (new SCIPION portal), CIRMMP, the University of Utrecht, and the NKI.

Sustainability

D1.4 set out a process for developing a sustainability plan. This will be a basis on which West-Life can plan for the long term an excellent and fit-for-purpose virtual service with the core activity being the provision of integrated tools and services through the VRE, and also allow for innovation and expansion to meet new demands.

5. Deviations from Annex 1 and Annex 2

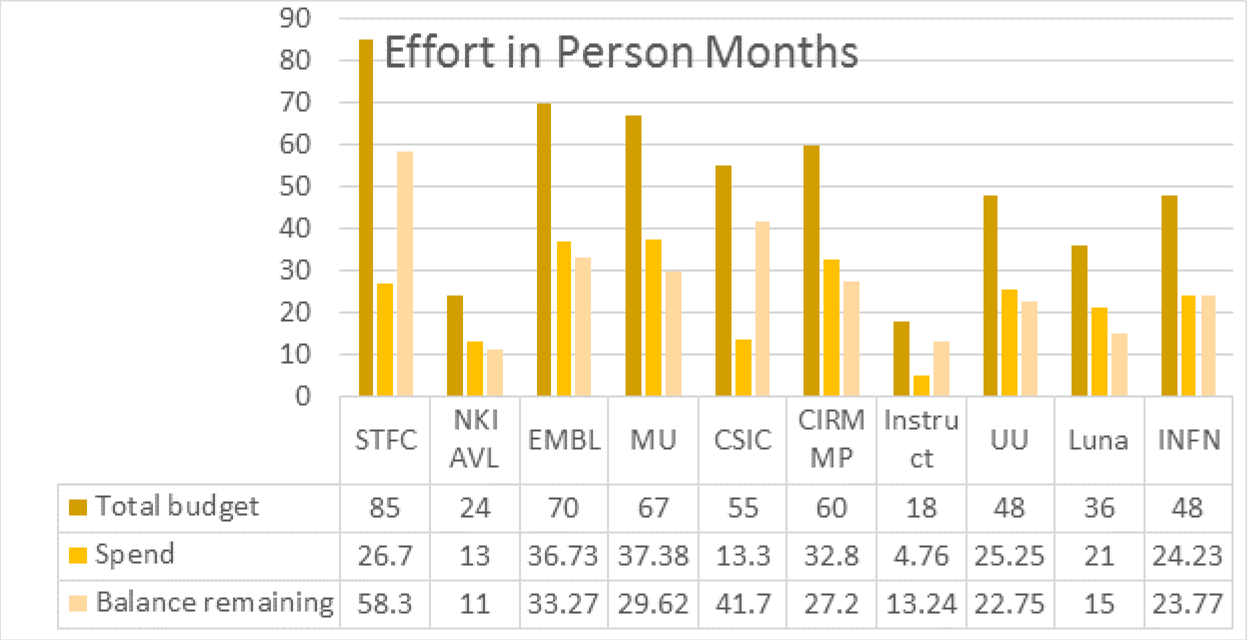
5.1 Tasks

D3.3 “Organisation of a round table or joint meeting involving ESFRIs” was held during the Instruct Biennial in May 2017, jointly with iNext, and writing of D3.2 “Report on the requirements of other infrastructures” is ongoing. It proved impractical to hold this meeting earlier. (This also led to delays to the related M7 and M9.)

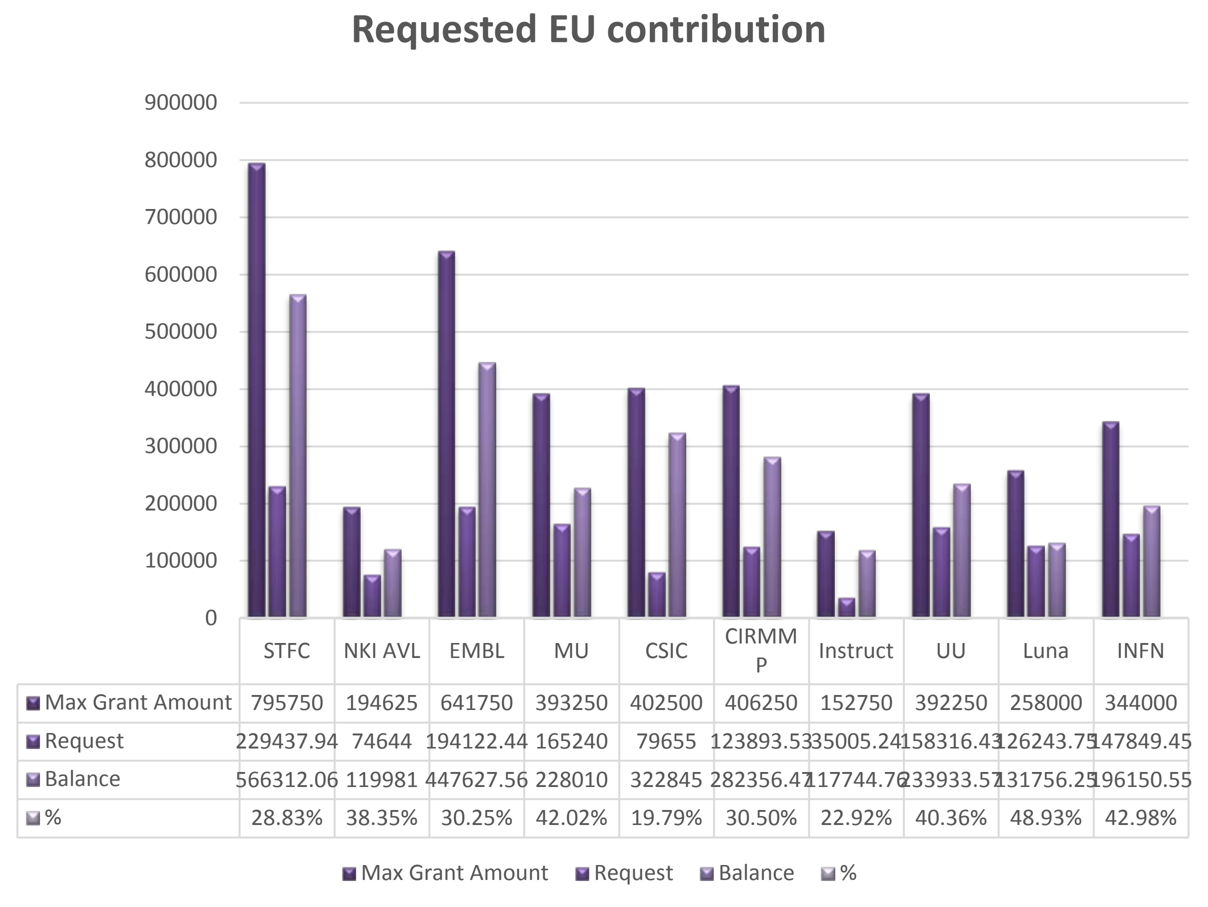
D5.4 “Report on the activities of the help desk” is also delayed. We are reviewing how best to complement the established user help resources in this field. The delay in the operational launch of the portal delayed the helpdesk planning. Identification and focused work on use cases also create the need for a targeted helpdesk approach. New staff recruitment and advancement of the portal implementation now allows us to engage with this as a priority, and we will deliver D5.4 in September.

5.2 Use of resources

During RP1, recruitment took longer than envisaged for some partners and personnel costs and effort were lower than expected. However, to date this has not affected completion of deliverables to the required deadlines and all partners have now recruited and plan to spend in line with the budget during RP2. CIRMMP have lower personnel costs than expected due to difficulties in recruiting at the start of the project and junior staff assisted with some of the initial work to ensure deliverables were met to time.



	STFC	NKI AVL	EMBL	MU	CSIC	CIRM MP	Instruct	UU	Luna	INFN
■ Total budget	85	24	70	67	55	60	18	48	36	48
■ Spend	26.7	13	36.73	37.38	13.3	32.8	4.76	25.25	21	24.23
■ Balance remaining	58.3	11	33.27	29.62	41.7	27.2	13.24	22.75	15	23.77



	STFC	NKI AVL	EMBL	MU	CSIC	CIRMMP	Instruct	UU	Luna	INFN
Max Grant Amount	795750	194625	641750	393250	402500	406250	152750	392250	258000	344000
Request	229437.94	74644	194122.44	165240	79655	123893.53	35005.24	158316.43	126243.75	147849.45
Balance	566312.06	119981	447627.56	228010	322845	282356.47	117744.76	233933.57	131756.25	196150.55
%	28.83%	38.35%	30.25%	42.02%	19.79%	30.50%	22.92%	40.36%	48.93%	42.98%

5.2.1 Unforeseen subcontracting

Not applicable.

5.2.2 Unforeseen use of in kind contribution from third party against payment or free of charges

The University of Florence has to be considered as a third party which makes available its resources to partner CIRMMP. The University of Florence and particularly CERM (Center of Magnetic Resonance) of the University of Florence (member of the CIRMMP) provides research contracts for PhD students and post docs. The Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP) reimburses the cost of such contracts to the University on the basis of specific agreements signed between CIRMMP and the University of Florence. The fellow(s) is/are required to carry out the research in specific educational/research programme with reference to the activities planned within the project. CIRMMP declares in its Form C these costs according to the General Model Grant Agreement (article 11). Estimated costs for this contribution is €110 000.

The University of Oxford is an in-kind third party which available its resources to partner Instruct, in accordance with Instruct Academic Services Limited’s role as subsidiary company of the University of Oxford. The University of Oxford provides employment contracts for staff seconded to Instruct. The seconded staff are required to carry out the activities planned within the project. The University of Oxford charges the staff cost by invoice to Instruct Academic Services Limited and Instruct declares these costs in its Form C according to the General Model Grant Agreement (Article 11). The in-kind contribution from the University of Oxford will be 18PM, €90 000