Deliverable D4.3

Project Title:	World-wide E-infrastructure for structural biology
Project Acronym:	West-Life
Grant agreement no.:	**675858**
Deliverable title:	Report on experience with deployment of consolidated platform and its interaction with infrastructure
WP No.	4
Lead Beneficiary:	4: Masarykova univerzita
WP Title	Operation and maintenance of the computing and data infrastructure
Contractual delivery date:	Month 15
Actual delivery date:	Month 15
WP leader:	Name: Ales Krenek — Partner: MU
Contributing partners:	STFC, EMBL-HH, MU, CSIC, CIRMMP, UU, Luna, INFN

Deliverable written by Marco Verlato, Ales Krenek, Alexandre Bonvin, Chris Morris, Antonio Rosato, Tomas Kulhanek

Contents

1 Executive summary

This document reports about the activities of the Work Package 4, which, building on a considerable legacy of existing operational web portals developed and used by the Structural Biology / WeNMR user community, aims at provisioning a consistent e-infrastructure to allow gradual integration of the existing isolated solutions to a single computing and data processing environment, based on the state of the art grid and cloud open source software tools and frameworks.

This report follows the documents D4.1 and D4.2, delivered at project month 9, and the document M4.2 delivered at month 12, so that only the progresses achieved until project month 15 not already described previously will be reported here.

The document starts with an updated description of the resources potentially available for the project from the EGI e-infrastructure, on top of which the consolidated West-Life platform is being built. It then presents a detailed view of resource usage and their geographical distribution in the first year of the project, as obtained from the EGI Accounting Portal. The remaining of the document reports the latest experiences and updates the activity plans on the consolidated job management mechanism, the programmatic access to datasets and the unified security and accounting model.

2 Project objectives

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

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

3 Detailed report on the deliverable

3.1 The HTC and Cloud production e-infrastructure

The consolidated platform of West-Life leverages the resources provided by the EGI. EGI is a federation of computing and storage resource providers united by a mission to support research and development. The EGI federated e-infrastructure is publicly funded (is supported by the EU project EGI-Engage) and provides compute and storage resources to support research and innovation. As of November 2016, the EGI federation comprises over 300 data centres, located mostly in European countries, and a number of integrated resource providers in Canada, USA, Latin America, Africa and the Asia-Pacific region. A total of 826,500 CPU-cores are available for the High Throughput Computing (HTC) platform, while 6,600 CPU-cores are available for the Cloud platform (known as the EGI Federated Cloud). Furthermore, 285 PB and 280 PB are respectively available as online and archive storage.

In the following sub-sections we summarize the use made by West-Life project of both HTC and Cloud e-infrastructures, in term of the metrics which were selected as KPI for the entire Work Package.

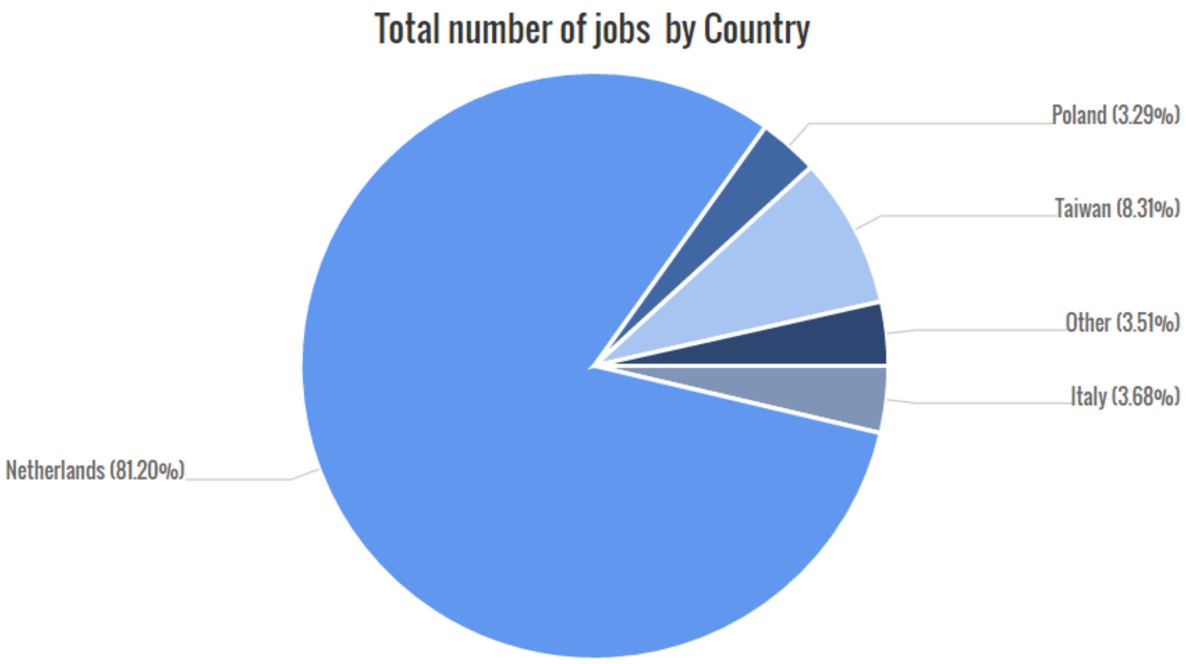
3.1.1 HTC e-infrastructure usage

West-Life users get access to the EGI HTC e-infrastructure (also known as EGI Grid) via the enmr.eu Virtual Organisation (VO). The enmr.eu VO was established in 2007 in the context of the EU project e-NMR and further supported in the follow-up EU project WeNMR. Out of the 826,500 CPU-cores provided in total by the EGI grid, around 211,000 belong to the 38 resource centres that support the enmr.eu VO. This support is generally shared with many other VOs, so that the effective availability of resources to the project is difficult to estimate, but can be argued by the EGI accounting system. The 38 resource centres are distributed in 13 countries around the world, and the pictures below show the total number of jobs and the normalized CPU time (in HEPSPEC06 hours) provided by each country in the first year of the West-Life project, from November 2015 until the end of October 2016. In 2016 EGI.eu and the MoBrain Competence Centre (a consortium represented by the Faculty of Science – Chemistry of Utrecht University) have signed a Service Level Agreement (SLA) which granted to enmr.eu VO for the whole year an amount of opportunistic computing time up to 55 Million of normalized CPU hours and opportunistic storage capacity up to 61 TB [1]. A total of seven resource centres signed the SLA: INFN-PADOVA (Italy), RAL-LCG2 (UK), TW-NCHC (Taiwan), SURFSara and NIKHEF (The Netherlands), NCG-INGRID-PT

(Portugal), CESNET-MetaCloud (Czech Republic). By signing the SLA the above resource providers committed to: operate 24x7x365 excluding planned maintenance windows or service interruptions which have to be notified via e-mail in timely manner (i.e. 24 hours before the outage); provide support through the EGI Service Helpdesk (http://helpdesk.egi.eu/), from Monday to Friday and from 9:00 to 17:00 in the time zone of the relevant Resource Centres; handle incidents according to the Quality of Support level that is estimated according to the impact of the outage or service quality degradation; target a minimum monthly availability (defined as the ability of a service or service component to fulfil its intended function at a specific time or over a calendar month) of 85% and monthly reliability (i.e. availability excluding scheduled maintenance periods) of 90%. According with the EGI accounting system more than 3.3 Million jobs have run translating into 19.6 Million of normalized CPU hours consumed by the West-Life applications. In both metrics we can notice a slight increase in the usage in the second semester, and the foremost contribution (above 80%) of Dutch resource centres (the Dutch NGI has a strong commitment to support life sciences). We should note here that jobs are not being targeted to specific sites and the distribution of jobs is left to the workload management system used (WMS and DIRAC4EGI).

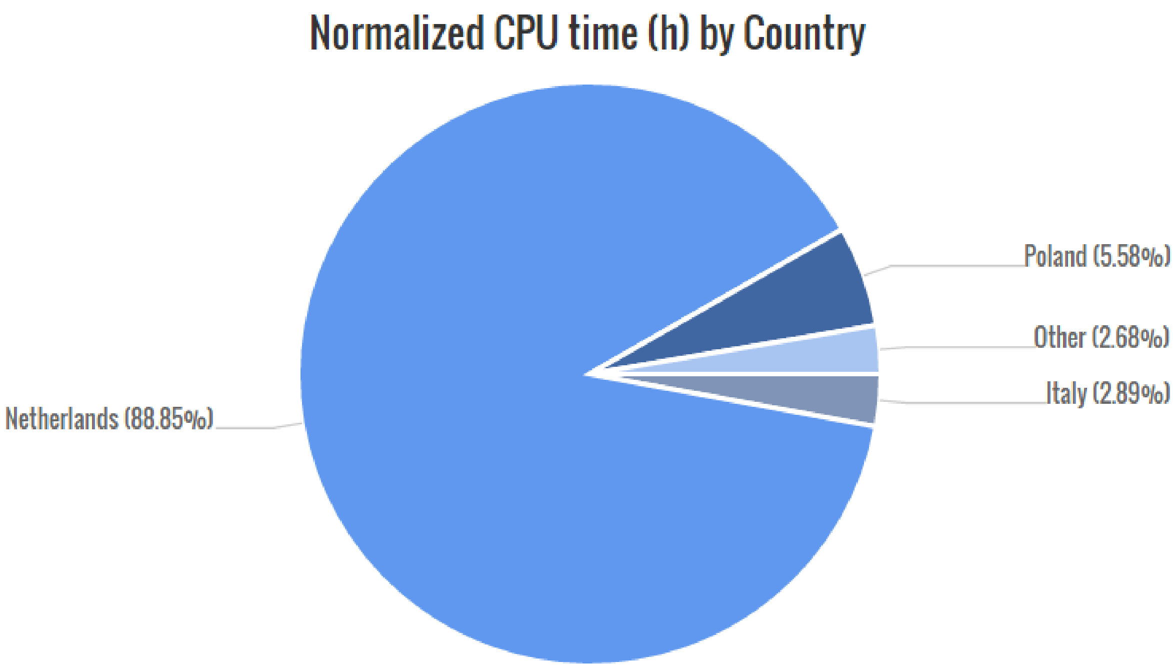
VO Admin - Total number of jobs by Country and Semester

Country	Nov 2015 - Apr 2016	May 2016 - Oct 2016	Total	Percent
Belgium	173	127	300	0.01%
Brazil	1,461	2,964	4,425	0.13%
China	580	477	1,057	0.03%
France	1,075	2,771	3,846	0.11%
Germany	16,834	24,191	41,025	1.21%
Italy	50,014	74,732	124,746	3.68%
Malaysia	4,655	945	5,600	0.17%
Netherlands	1,339,680	1,409,443	2,749,123	81.20%
Poland	47,957	63,556	111,513	3.29%
Portugal	24,345	33,056	57,401	1.70%
South Africa	187	22	209	0.01%
Taiwan	147,341	133,843	281,184	8.31%
United Kingdom	928	4,100	5,028	0.15%
Total	1,635,230	1,750,227	3,385,457	
Percent	48.30%	51.70%		



VO Admin - Normalized CPU time (h) by Country and Semester

Country	Nov 2015 - Apr 2016	May 2016 - Oct 2016	Total	Percent
Belgium	1,834	5	1,839	0.01%
Brazil	24	36	60	0.00%
China	0	211	211	0.00%
France	139	137	277	0.00%
Germany	46,925	138,543	185,468	0.94%
Italy	274,420	293,188	567,608	2.89%
Malaysia	1,844	87	1,931	0.01%
Netherlands	7,343,769	10,130,521	17,474,290	88.85%
Poland	414,972	683,197	1,098,169	5.58%
Portugal	88,953	208,999	297,952	1.51%
South Africa	0	0	0	0.00%
Taiwan	12,297	27,534	39,832	0.20%
United Kingdom	49	41	90	0.00%
Total	8,185,227	11,482,498	19,667,725	
Percent	41.62%	58.38%		



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 DisVis and PowerFit applications to greatly enhance their performance. Two new web portals making using of those GPGPU resources have now been put into operation (http://milou.science.uu.nl/enmr/services/DISVIS and http://milou.science.uu.nl/enmr/services/POWERFIT) [2] . These make use for this purpose of GPGPU-enabled Docker containers (https://hub.docker.com/u/indigodatacloudapps/) developed in collaboration with the INDIGO-DataCloud project (see also section 3.2.1.2). GPGPU support was transparently integrated also in the AMPS-NMR (AMBER) portal.

3.1.2 Cloud e-infrastructure usage

The enmr.eu VO is also the door for accessing the EGI Federated Cloud. In this case, out of the 6,600 CPU-cores provided in total, 624 belong to the 3 resource centres that support the enmr.eu VO: these are CESNET-Metacloud in Czech Republic, INFN-PADOVA-STACK in Italy and IISAS-GPUCloud in Slovakia. The pictures below show the total number of Virtual

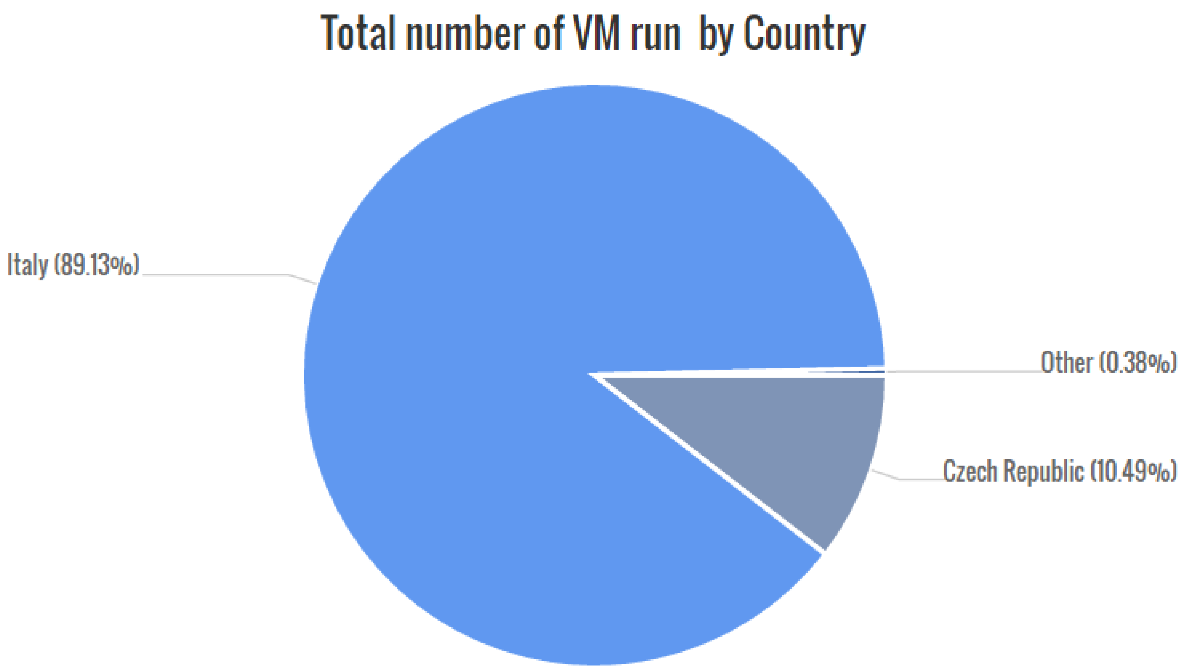
Machines (VM) and their elapsed time (in hours) provided by each country in the first year of the West-Life project, from November 2015 until the end of October 2016. More than 1,300 VMs and 151,000 CPU wall time hours have been consumed by the West-Life applications (around 45% by Scipion, the rest by Haddock, DisVis, PowerFit and Gromacs).

The CESNET-Metacloud and IISAS-GPUCloud sites also made available respectively 4 NVIDIA Tesla M2090 and 4 NVIDIA Tesla K20m GPU cards, giving therefore the possibility to instantiate VMs with one or more GPU devices attached via PCI passthrough virtualisation, a solution implemented by IISAS in the context of the EGI-Engage project.

The STFC is working on making its Tier 1 compute resource available for West-Life jobs. The preferred method is to implement a site fully compliant with the EGI Federated Cloud facility, but depends on the release of the new CloudKeeper functionality. If there are further delays then STFC will implement support for DIRAC job submission in advance of other Federated Cloud functionality.

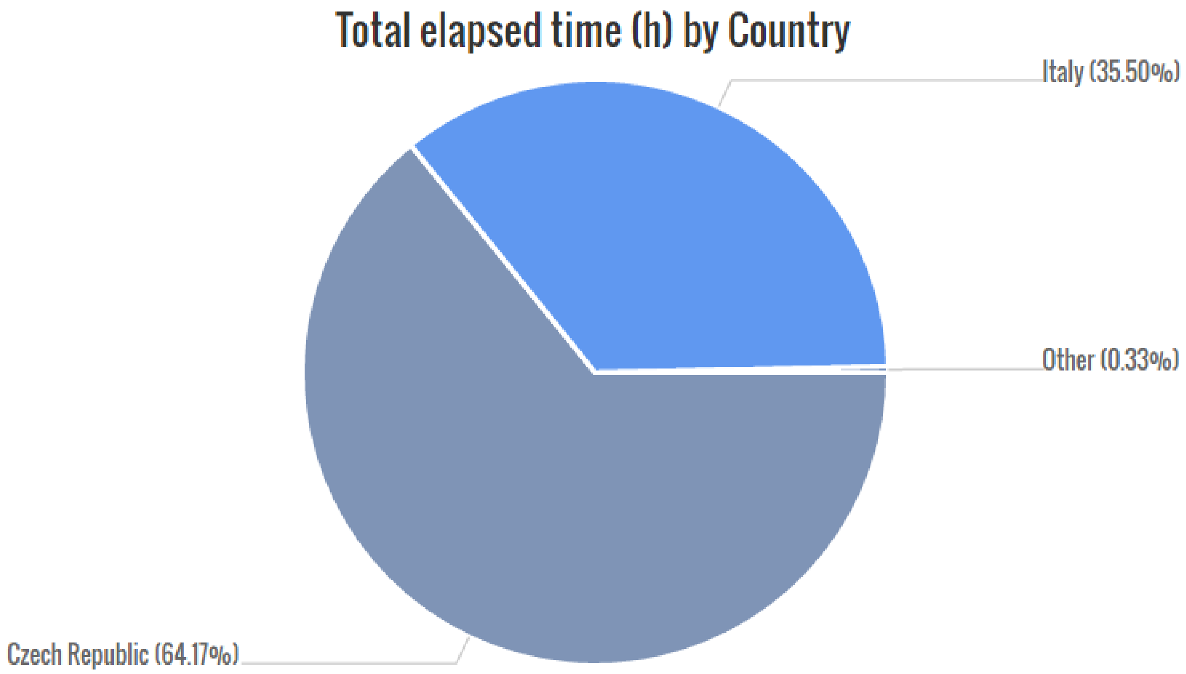
Cloud VO Admin - Total number of VM run by Country and Semester

Country	Nov 2015 - Apr 2016	May 2016 - Oct 2016	Total	Percent
Czech Republic	61	78	139	10.49%
Italy	516	665	1,181	89.13%
Slovakia	0	5	5	0.38%
Total	577	748	1,325	
Percent	43.55%	56.45%		



Cloud VO Admin - Total elapsed time (h) by Country and Semester

Country	Nov 2015 - Apr 2016	May 2016 - Oct 2016	Total	Percent
Czech Republic	68,043	29,052	97,096	64.17%
Italy	17,562	36,150	53,712	35.50%
Slovakia	0	498	498	0.33%
Total	85,605	65,701	151,306	
Percent	56.58%	43.42%		



3.2 The consolidated platform on top of the e-infrastructure

3.2.1 Consolidation of job management mechanism

The common job management mechanism behind the application portals, which distribute the highly demanding payloads across the HTC e-infrastructure, has been widely described both in Deliverable 4.1 delivered at month 9, and in Milestone 4.2 delivered at month 12. Accordingly, here we will only report the latest updates.

3.2.1.1 DIRAC4EGI service

As pointed out in the D4.1 and M4.2 documents, DIRAC (Distributed Infrastructure with Remote Agent Control, http://diracgrid.org/) was the software chosen to implement the West-Life job dispatcher. Its main advantage is the capability to submit jobs to both HTC and Cloud available resources of EGI, through a uniform API exposed to the users/portal developers. Moreover, DIRAC developers in conjunction with EGI operate and maintain since a couple of years the DIRAC4EGI service, a cluster of instances to allow any user belonging to an EGI supported VO to distribute computational tasks among the available EGI resources, monitor their status and retrieve the results.

So far only the HADDOCK portal has used DIRAC4EGI in production mode to access the EGI HTC resources. However, in the last months the UU team has investigated the ability of DIRAC to launch VMs on the EGI Federated Cloud. This scenario has been successfully demonstrated for the Gromacs application with help of one of the DIRAC4EGI developer, Dr.

Victor Mendez (UAB, Spain). We plan to migrate the Gromacs grid-enabled portal to a DIRAC submission mechanism to access multi-core VMs of the Federated Cloud.

In order to increase the DIRAC4EGI awareness within project partners, a workshop was held, collocated with the West-Life All Partners Meeting in January 2017, to show portal developers how to convert legacy grid portals to DIRAC.

3.2.1.2 Application Software Deployment

This topic has also been widely discussed in the D4.1 and M4.2 documents. In the HTC e-infrastructure the sites supporting a given application are advertised in the grid information system (BDII) and the Workload Management System (gLite-WMS or DIRAC) is able to target them by matching the job requirements specified in the Job Description Language (JDL). Application software packages are made available at the site through the CernVM File System (CVMFS), a scalable, reliable and low-maintenance software distribution service. CVMFS was developed at CERN to assist High Energy Physics (HEP) collaborations to deploy software on the worldwide-distributed computing infrastructure used to run data processing applications. In the work described in the previous sub-section CVMFS has been demonstrated to be also a good mechanism to provision software on the VMs launched by the DIRAC service.

The current production software repository used by services of the enmr.eu VO on HTC sites is the WeNMR repository wenmr.egi.eu. This repository can be accessed currently without any restriction at all sites supporting the enmr.eu VO. This is also the repository that is mounted in the DIRAC VMs.

Under the recent development a new CVMFS repository was established to hold and distribute the software which is not yet distributed with existing repositories. The CVMFS repository is accessible at west-life.egi.eu and is currently available at 3 stratum 1 sites mirroring the content.

Scipion version 1.0.1 and Scipion web tools 1.0.1 are available under /cvmfs/west-life.egi.eu/software/scipion. The Virtuoso-opensource server combining relational, graph and document data management is distributed under /cvmfs/west-life.egi.eu/software/virtuoso.

The CCP4 software suite is distributed in another CVMFS repository facilities.gridpp.ac.uk maintained by the CCP4 development team. This repository can be accessed after the license is agreed by an end user and supplying an X.509 key.

In the EGI Federated Cloud application software is often embedded in the Virtual Machine Images (VMI) published in the EGI Application Database (AppDB). Currently two of such

VMIs, also called Virtual Appliances, have been produced by West-Life partners and published in the image list of enmr.eu VO:

- Westlife WP6 Virtual Folder (https://appdb.egi.eu/store/vappliance/d6.1.virtualfoldervm) is a VM based on CernVM system booting to Scientific Linux 7.2. It mounts remote installation of software at cernvm.ch and west-life.egi.eu. CernVM uses similar approach as CVMFS to distribute the operating system. An initial small image (18MB) with bootloader and contextualization configures and prepares system with full operating system capabilities (currently Scientific Linux – which is open source clone of RHEL maintained by FermiLab). The CernVM was tested with distributing the virtual folder functionality, see D6.1 document for further details;
- ScipionCloud_v1.0 (https://appdb.egi.eu/store/vappliance/scipion.v1.0) is an image processing framework for obtaining 3D models of macromolecular complexes using Electron Microscopy (3DEM). This image provides Scipion v1.0 to be run on a single node, plus remote desktop guacamole installed and configured to access the Virtual Machine through a Web Browser.

Recent developments leverage the user-space restricted Docker environment Udocker (https://www.indigo-datacloud.eu/userspace-container-support). This tool enables container deployment without superuser privileges, hence it can be used in both legacy grid jobs and virtual machines. In particular, the collaboration with INDIGO-DataCloud project has allowed us to embed the AmberTools, DisVis and PowerFit application software (including all their dependencies like e.g. FFTW3, pyFFT and OpenCL libraries) in Docker images based on Ubuntu, which are now maintained by the INDIGO-DataCloud project and distributed via dockerhub at their repository https://hub.docker.com/u/indigodatacloudapps/.

3.2.3 Deployment of applications in the Cloud e-infrastructure

The portal virtualization activity and the experimental work to set up DisVis, PowerFit and Scipion applications directly in the EGI Federated Cloud was already described in the M4.2 document. We will report here the progress achieved for other applications.

3.2.3.1 AMPS-NMR

As also described in D4.1, the CIRMMP partner has been evaluating the use of one of the solutions developed by the INDIGO-DataCloud project, namely the FutureGateway API, to submit calculations to the HTC or Cloud environments transparently. The potential advantage of using the FutureGateway is its readier integration with cloud storage solutions such as OneData, another development of INDIGO-DataCloud. The plan is to use the entire set-up of INDIGO-DataCloud, from authentication, using the IAM service, to job submission and cloud storage, by embedding them behind the already existing AMPS-NMR interface. Thus, similarly to the DisVis and PowerFit servers, the application is virtualized by leveraging suitably prepared Docker containers (see 3.2.1.2).

At present, the integration of the IAM and FutureGateway services in the AMPS-NMR production portal has been completed. Work is in progress to further integrate OneData cloud storage (see also next section).

3.2.3.2 Gromacs

Development of automated cloud deployment of the whole Gromacs portal continued from the prototype described in M4.2. Specifically, restrictions on number of worker nodes in the deployment were worked around, which is the first step to dynamic deployments (regulating the number of allocated resources according to the load of the portal). The whole sequence of molecular dynamics protocol was ported to Gromacs version 5. The deployment process detects architecture of the worker node CPU, and chooses from specifically optimized Gromacs binaries. GPU support is also available. Thorough implementation of the reverse process – cleanup and deallocation of resources was added. Deployment of the Gromacs 5 West-Life portal will be done using this mechanism, and it will be made available to a wider user community in the upcoming period.

3.2.3.3 Scipion Web Tools

Following the pattern of Gromacs deployment (using TOSCA description and Cloudify orchestrator), the first prototype of Scipion Web Tools was sketched, and the specific application dependent problems are being addressed (in particular, the need for large semi-persistent storage).

3.2.4 Programmatic access to datasets

Work Package 4 has a task aiming at defining an architecture and the appropriate interfaces to access the relevant biological datasets, including strategies to cache copies of the data across the distributed e-infrastructure. This activity was planned to start in the second year of the project, building on the experience and the intermediate output of the Work Package 6, which was designed to provide application layer services for data management suitable for the growing use of multiple techniques and multiple experimental facilities in structural biology research projects. WP6 intermediate results have been timely delivered at month 12 in the document D6.1: Virtual Folder, and will not be described here.

In November 2016 INFN has started the deployment of OneData, a component of the Unified Data Access (UDA) layer released by INDIGO-DataCloud project in August 2016. As

described in deliverable D4.1, it is designed to support efficient, high performance and scalable distributed data caching mechanisms, integration with existing storage infrastructures, and transparent client-side import/export of distributed cloud data.
INFN has procured a Dell MD3600 storage server with 12 x 2TB disks, which is being installed with OneData software at INFN-PADOVA-STACK site of the EGI Federated Cloud. The initial goal is to use it for testing a new West-Life application under development by the CIRMMP partner.

3.3 Unified security and accounting model

The D4.2 document delivered at month 9 had reviewed the existing Authentication and Authorisation Infrastructures (AAI) used in West-Life and provided an updated architecture, based on the current best practices and state of the art technology, aimed at harmonising the different approaches taken by existing components in order to achieve a single AAI used through the project. The M4.2 document described the first implementation of AAI components following the designed architecture, focusing in particular on the prototype of the IdP-SP-Proxy service deployed in the testbed.

After finalizing M4.2, an agreement was reached to connect the IdP-SP proxy instance with the existing WeNMR SSO service, and technical details were clarified. This step will allow smooth migration of the WeNMR users to new West-life services (Gromacs version 5 being the pilot) while allowing to move to other identity providers (ARIA, ORCID, etc.) gradually.

References cited

[1] VO SLA MoBrain and VO OLAs MoBrain, https://documents.egi.eu/document/2751

[2] G.C.P. van Zundert, M. Trellet, J. Schaarschmidt, Z. Kurkcuoglu, M. David, M. Verlato, A. Rosato and A.M.J.J. Bonvin. The DisVis and PowerFit web servers: Explorative and Integrative Modeling of Biomolecular Complexes. J. Mol. Biol., (2017) **429**, 399-407

Background information

This deliverable relates to WP4; background information on this WP as originally indicated in the description of work (DOW) is included below.

WP4 **Title:** Operation and maintenance of the computing and data infrastructure
Lead: MU
Participants: STFC, EMBL, MU, CSIC, CIRMMP, UU, Luna, INFN

Work package number	4	**Start date or starting event:**	1
Work package title	Operation and maintenance of the computing and data infrastructure		
Activity Type	OTHER		
Participant number			
Person-months per participant:			

Objectives
The principal objectives of this work package are:
- **O4.1:** Setup the project testbed, define interfaces used to provision hardware resources, and negotiate provisioning with the resource providers at the technical level
- **O4.2:** Define, implement, and deploy consolidated architecture for job submission and data access.
- **O4.3:** Review existing security frameworks and define consolidated solution
- **O4.4:** Ensure smooth migration of the legacy portals to the consolidated architecture.

Description of work and role of participants
Task 4.1: Consolidation and operation of the infrastructure
Leader: INFN
Participants: MU, STFC, EMBL-HH, UU, CIRMMP, LUNA
The task is responsible to continue and improve the operations of previously developed computational platforms and services, including those developed within the previous WeNMR project as well as the ones already offered by the different partners addressing issues in X-ray crystallography and cryo-EM. We will achieve a better integration among them and simplify user interaction over a large set of experimental techniques in Structural Biology.
The West-Life project will leverage for resource provisioning both the EGI High-Throughput Data Analysis platform (the grid) and the EGI Federated Cloud, and possibly other public commercial or private cloud providers. The inventory of the computing and storage resources supporting the project and the selection of services available from the existing production e-infrastructures and/or EU projects to access, manage and operate these resources in a coherent manner will be done at the beginning of the project (M4.1), including a testbed available for rapid prototyping. The testbed will become the basis for the West-Life production platform. Because the introduction of new features and migration to new interfaces must not jeopardise the scientific work of the current user communities, we expect this platform to evolve in several generations that will co-exist in time in the phases of

unstable integration, stable production, and eventual phase-out. Times of switching an integrated next generation of the platform to production are the principal milestones of the work package (M4.2, M4.4, and M4.5).
All the partners are either directly involved or they collaborate closely with their National Grid Infrastructures (NGIs), hence having direct connection to EGI. Those synergies will be leveraged during operation of the project infrastructure rather than building it independently. In particular, established operational mechanisms for infrastructure monitoring, reliability evaluation, security alert and incident management etc. will be reused.

Task 4.2: Consolidation of job management mechanisms

Leader: LUNA

Participants: MU, INFN, UU, CIRMMP

An important point of the entire architecture is the interaction with the underlying cloud e-Infrastructure. Due to the legacy of current solutions, many different dispatchers are used submitting jobs to various infrastructures. This approach brings non-negligible overhead of maintaining these components, and it is not sustainable in the long term. On the contrary, uniform dispatcher technologies compatible with cloud interfaces have emerged over the past few years (DIRAC, HTCondor, Mesos, and Docker -container technology-). In addition, standardized interfaces to cloud resources running various cloud-management middleware (OpenStack, OpenNebula, ...) and even commercial clouds (Amazon EC2, Microsoft Azure) will be considered. In collaboration with T4.1, gradual migration from legacy interfaces to the standardized ones will be negotiated with resource providers so that legacy interfaces can be phased out smoothly, without affecting the user community.
Specific scientific code is expected to be wrapped in virtual machines in a standardized way so that further applications can be added easily. This task will develop guidelines for such wrapping in order to integrate with the dispatcher smoothly (preparation of specific VM images is done elsewhere, WP5 and WP7 in particular). For this, the project will select a consistent solution in its initial phase after a detailed analysis, and a consolidated architecture will be proposed in D4.1.
The interfaces among the components of the entire architecture are critical for its stable operation. While the work of WP5 is development of the Web Portal/VRE, it’s the responsibility of WP4 to specify the interfaces/API precisely, so that the VRE can connect to the dispatcher in a seamless and reliable way. The interface specifications are included in the architecture description deliverables D4.1, and they will be revised according to deployment experience in D4.5 and D4.6.
Further, the present task will keep supporting deployment of the portal frontend and applications using the consolidated job management mechanism. Based on experience with the deployment and according to incoming detailed requirements, the job submission architecture will be updated and revised eventually (thus contributing to deliverables D4.3, D4.5, and D4.6).

Task 4.3: Programmatic Access to datasets

Leader: STFC

Participants: LUNA, INFN

Many services rely on existing external data sets. The infrastructure will have to be capable of leveraging this external data upon users’ request. The task will review the relevant datasets to be made available, and it will define architecture and appropriate interfaces to access them, including eventual strategies to make “caching” copies of the data. It will be built on the metadata services to be offered in WP6.
Together with T4.4 security issues (authorization and user identity delegation in particular) will be addressed.
Unlike the other tasks of WP4, T4.3 brings functionality that is not widely present in current portal solutions. Therefore, it starts later (year 2) in order to build on experience and intermediate outcomes of WP6.

Task 4.4: Unified security and accounting model

Leader: MU

Participants: INFN, LUNA
e-Infrastructures require the users of the precious resources (computing power and storage capacity) to be reliably authenticated. Typically, rather heavyweight mechanisms like X.509 certificates are used. On the other hand, when the user interacts with the infrastructure through an application portal, lightweight mechanisms (username + password) are strongly preferred. Identity federations, which allow the user to authenticate with his/her home institute credentials, have gained popularity recently.
Due to various technical reasons, the primary user credentials are normally not passed or mapped directly to the computing and storage resources. Instead, portals often use “robot” certificates – the identity of the portal itself, which is accepted by the resources. Users are identified with a lightweight mechanism with the portal, and the portal must maintain mapping of the users credentials to specific use of the infrastructure with the robot credential in order to ensure accountability and traceability.
The task will review security mechanisms used by the various community portals, and it will design a unified mechanism for the project, balancing the impact on the existing users with clean design and maintainability of the resulting solution. The design will be specified in D4.2, and its gradual implementation will be part of the consolidated deployments – milestones M4.2, M4.4, and M4.5.

Deliverables

No.	Name	Due month
D4.1	Consolidated architecture of job submission and interaction with infrastructure	9
D4.2	Common security model design	9
D4.3	Report on experience with deployment of consolidated platform and its interaction with infrastructure	15
D4.4	Overview of external datasets, strategy of access methods, and implications on the portal architecture	18
D4.5	Report on progress of the deployment of consolidated platform and its interaction with infrastructure	26
D4.6	Final report on deployment of consolidated platform and the overall architecture	36