Deliverable D5.7

Project Title:	World-wide E-infrastructure for structural biology	
Project Acronym:	West-Life	
Grant agreement no.:	**675858**	
Deliverable title:	Report on access and usage statistics of the various services	
WP No.	5	
Lead Beneficiary:	8: Utrecht University	
WP Title	Virtual Research Environment	
Contractual delivery date:	24	
Actual delivery date:	24	
WP leader:	Alexandre M.J.J. Bonvin	Utrecht University
Contributing partners:	STFC, NKI, EMBL, MU, CSIC, CIRMMP, Instruct, UU, Luna, INFN	

COPYRIGHT NOTICE

DELIVERY SLIP

	Name	***Partner/Activity***	***Date***
From:	J. Schaarschmidt / A. Bonvin	UU/WP5	27.10.2017
Approved by:			

DOCUMENT LOG

Issue	***Date***	***Comment***	***Author/Partner***
v.1	17.10.2017	First version	J. Schaarschmidt/A.Bonvin / UU
v.2	27.10.2017	Second version	J. Schaarschmidt/A.Bonvin / UU

Contents

1

Executive summary

— West-Life aims to bring complex data analysis in Structural Biology to a simple Web browser-based Virtual Research Environment (VRE). Capitalizing on European and National projects, such as Instruct, WeNMR, and CCP4, as well as other projects, like EGI-Engage, through the MoBrain Competence Center, West-Life thus starts from a leading position in the Structural Biology field, in that its partners are already providing valued services in specific disciplines. After defining the baseline performance of 19 pre-existing web services provided by 6 project partners in Deliverable 5.2, and reporting an update in the Periodic Report, we now provide usage statistics for 2017 for all services in operation at this date. These show a sustained growth of the majority of services as well as an increasing usage of the newly introduced services. Furthermore, we introduce five new services (3 of which were described in the mid-term periodic report) - 3DBionotes , CCP4-MoRDa, PowerFit, DisVis, and SpotON - which were recently made available to the scientific community through the support of the West-Life project

2 Project objectives

This deliverable is contributing 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 Background

The overarching objective of the West-Life 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. Capitalizing on European and National projects, such as Instruct, WeNMR, and CCP4, as well as other projects, like EGI-Engage, through the MoBrain Competence Center formed by several West-Life partners, a series of Web Services addressing specific pipelines in NMR, X-ray diffraction, SAXS and cryo Electron Microscopy data analysis are offered with direct impact on a large and worldwide user base. West-Life has by now been operating for almost two years web services for the Structural Biology community. In the following, an update is given on the use and adoption of the 23 portals operated by West-Life partners. Five of those are newly developed portals added to the West-Life portfolio.

3.2 Description of new portals

The pre-existing services provided by the West-Life partners have been described in Deliverable 5.2. Here we briefly summarize the new services introduced since the publication of Deliverable 5.2 a year ago. These consist of three services already included in the Periodic Report for the mid-term review (DisVis, PowerFit, and MoRDa) and two new services (SpotOn and 3DBionotes).

3.2.1 DisVis

DisVis allows 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.

3.2.2 PowerFit

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.

3.2.3 CCP4 - MoRDa

MoRDa is an alternative to two other CCP4online services 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 on-going to provide REST APIs and integration with the Virtual Folder for all CCP4 online portals. West-Life has helped to link the output of MoRDa to ARP/wARP and PDB_REDO.

3.2.4 SpotON

SpotOn is a robust algorithm developed to identify and classify the interfacial residues as Hot-Spots (HS) and Null-Spots (NS) with a final accuracy of 0.95 and a sensitivity of 0.95 on an independent test set. The predictor was developed using an ensemble learning algorithm with up-sampling of the minor class and was trained on a large number of complexes and on a high number of different structural- and evolutionary sequence-based features.

3.2.5 3DBionotes

3DBionotes is a web application designed to automatically annotate biochemical and biomedical information onto structural models. Current sources of information include post-translational modifications, genomic variations associated to diseases, short linear motifs, immune epitopes sites, disordered regions and domain families. 3DBionotes has also been integrated in the output page of other West-Life services such as HADDOCK.

3.3 Portal Statistics

Below we report usage statistics of the West-Life associated services including, where applicable, the number of users and job submissions for 2016 and 2017 respectively and their development in comparison to the previous reporting periods. KPIs are presented both for the 2016 (compared to the baseline of 2015) and 2017 (compared to the 2016 numbers). New services are highlighted by a green background. In general, the majority of portals show a sustained growth. We can also observe a good adoption and increasing usage of the newly introduced services.

The XPLOR server allows user to generate structural models from NMR data alone. Due to low demand for this service, after the end of the WeNMR project we postponed the update of the server und thus no statistics are reported for it. We recently completed this update, so that the server is again capable of sending jobs to the grid. We are now testing the server, and will identify new features to be introduced. Before the end of the year, we will re-open the server to users.

3.3.1 Portal statistics 2016

	Stats 2016				KPIs 2016			
	Users total	Users 2016	Job submissions	Grid/cloud jobs	%increase #Users total	%increase #Users 2016	%increase #Job submissions	%increase #Grid/cloud jobs
AMPS-NMR	390	90	185	3655	30%	80%	n.a.	-54%
ARP/wARP	4418	666	3466	0	8%	57%	7%	n.a.
Auto-Rickshaw	2550	430	4357	0	10%	-6%	22%	n.a.
CCD	0	0	4000	0	n.a.	n.a.	33%	n.a.
CCP4-AMPLE	187	187	535	0	n.a.	n.a.	n.a.	n.a.
CCP4-BALBES	1521	700	3148	0	32%	-4%	-5%	n.a.
CCP4-Crank2	94	94	587	0	n.a.	n.a.	n.a.	n.a.
CCP4-MORDA	202	202	790	0	n.a.	n.a.	n.a.	n.a.
CCP4-MRBUMP	874	453	1721	0	51%	-4%	32%	n.a.
CCP4-SHELX	151	151	656	0	n.a.	n.a.	n.a.	n.a.
CCP4-ZANUDA	390	196	598	0	47%	4%	61%	n.a.
CS-ROSETTA3	228	9	33	263871	4%	-44%	-51%	40%
DisVis	39	38	172	64	n.a.	n.a.	n.a.	n.a.
FANTEN				0	n.a.	n.a.	n.a.	n.a.
GROMACS	121	9	107	368	8%	-65%	-38%	-37%
HADDOCK	8320	1621	27291	8947942	24%	12%	10%	19%
PDB_REDO	1200	500	4400	0	71%	0%	47%	n.a.
PowerFit	21	21	34	13	n.a.	n.a.	n.a.	n.a.
SCIPION	0	0	418	30446	n.a.	n.a.	318%	1347%
UNIO	80	21	72	280	36%	-32%	16%	-88%
ViCi	417	131	196	0	43%	42%	111%	n.a.
	Users total	Users 2016	Job submissions	Grid/cloud jobs	%increase #Users total	%increase #Users 2016	%increase #Job submissions	%increase #Grid/cloud jobs
TOTAL	21203	5519	52766	9246639	30%	3%	43%	204%

3.3.2 Portal statistics 2017 (projected)

Statistics for 2017 were gathered in October and extrapolated to the full year. The numbers are therefore subject to change in subsequent reports.

	Stats 2017 (extrapolated)				KPIs 2017			
	Users total	Users 2017	Job submissions	Grid/cloud jobs	%increase #Users total	%increase #Users 2017	%increase #Job submissions	%increase #Grid/cloud jobs
3DBionotes	1406	1406	6954.	n.a.	n.a.	n.a.	n.a.	n.a.
AMPS-NMR	468	78	454	1383	20%	-13%	145%	-62%
ARP/wARP	4974	894	5362	0	13%	34%	55%	n.a.
Auto-Rickshaw	2893	461	4517	0	13%	7%	4%	n.a.
CCD	0	0	4412	0	n.a.	n.a.	10%	n.a.
CCP4-AMPLE	392	205	708	0	110%	10%	32%	n.a.
CCP4-BALBES	2344	823	3800	0	54%	18%	21%	n.a.
CCP4-Crank2	206	112	601	0	119%	19%	2%	n.a.
CCP4-MORDA	659	457	1775	0	226%	126%	125%	n.a.
CCP4-MRBUMP	1383	509	1837	0	58%	12%	7%	n.a.
CCP4-SHELX	285	134	645	0	89%	-11%	-2%	n.a.
CCP4-ZANUDA	612	222	554	0	57%	13%	-7%	n.a.
CS-ROSETTA3	228	0	20	817	0%	n.a	-39%	-100%
DisVis	160	121	553	218	310%	218%	222%	241%
FANTEN	n.a.	n.a.	1246	0	n.a.	n.a.	n.a.	n.a.
GROMACS	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
HADDOCK	10190	1870	33300	8893000	22%	15%	22%	-1%
PDB_REDO		468	4947	0	n.a	-6%	12%	n.a.
PowerFit	99	78	150	96	371%	271%	341%	638%
Scipion	n.a.	n.a.	443	40170	n.a.	n.a.	6%	32%
SpotOn	105	105	165	0	n.a.	n.a.	n.a.	n.a.
UNIO	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
ViCi	482	68	98	0	16%	-48%	-50%	n.a.
	Users total	Users 2017	Job submissions	Grid/cloud jobs	%increase #Users total	%increase #Users 2017	%increase #Job submissions	%increase #Grid/cloud jobs
TOTAL	25480	6605	65587	8935684	99%	44%	50%	125%

Due to a server failure, three services (UNIO, GROMACS, and CS-ROSETTA3 – highlighted in red) ceased operation in 2017. Of those CS-ROSETTA3 has been put back in service. For GROMACS and UNIO the submission machinery has to be updated and migrated to the consolidated submission machinery (see Milestone document MS14) and work is ongoing to implement this. However due to the reduced interest in UNIO as apparent from the 2016 statistics, this service might be discontinued in the future.

Appendix 1: Portal summary – DisVis

Portal name	**DisVis**
Short description	DisVis allows 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
Keywords	Integrative modelling; biomolecular complexes; HotSpot prediction
URL	http://milou.science.uu.nl/enmr/services/DISVIS/
Grid-enabled	yes
Cloud-enabled	no
Operational since	August 1st 2016
Total number of registered users (October 1st 2017)	137
Number of user submissions processed (October 1st 2017)	636
Number of grid/cloud jobs	258
Key references	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.*, Advanced Online Publication (2016). G.C.P. van Zundert and A.M.J.J. Bonvin (2015) DisVis: Quantifying and visualizing accessible interaction space of distance-restrained biomolecular complexes.." *Bioinformatics* **31**, 3222-3224.

Appendix 2: Portal summary – PowerFit

Portal name	**PowerFit**
Short description	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.
Keywords	Integrative modelling; cryoEM
URL	http://milou.science.uu.nl/enmr/services/POWERFIT/
Grid-enabled	yes
Cloud-enabled	no
Operational since	August 1st 2016
Total number of registered users (October 1st 2017)	85
Number of user submissions processed (October 1st 2017)	153
Number of grid/cloud jobs	89
Key references	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.*, Advanced Online Publication (2016). G.C.P. van Zundert and A.M.J.J. Bonvin (2015) Fast and sensitive rigid-body fitting into cryo-EM density maps with PowerFit." *AIMS Biophysics* **2**, 73-87.

Appendix 3: Portal summary – SpotON

Portal name	**SpotOn**
Short description	SpotOn is a robust algorithm developed to identify and classify the interfacial residues as Hot-Spots (HS) and Null-Spots (NS) with a final accuracy of 0.95 and a sensitivity of 0.95 on an independent test set. The predictor was developed using an ensemble learning algorithm with up-sampling of the minor class and was trained on a large number of complexes and on a high number of different structural- and evolutionary sequence-based features.
Keywords	Integrative modelling; biomolecular complexes; HotSpot prediction
URL	http://scipion.cnb.csic.es/m/services/
Grid-enabled	no
Cloud-enabled	no
Operational since	March 1st 2017
Total number of registered users (October 1st 2017)	73
Number of user submissions processed (October 1st 2017)	115
Number of grid/cloud jobs	Not applicable
Key references	Moreira and Koukos et al. (2017) SpotOn: High Accuracy Identification of Protein-Protein Interface Hot-Spots. *Scientific Reports* **7:**8007. Melo et al (2016) A Machine Learning Approach for Hot-Spot Detection at Protein-Protein Interfaces. *Int. J. Mol. Sci.* **17**, 1215.

Appendix 4: Portal summary – 3DBionotes

Portal name	**3DBionotes**
Short description	3DBionotes is a web application designed to automatically annotate biochemical and biomedical information onto structural models. Current sources of information include post-translational modifications, genomic variations associated to diseases, short linear motifs, immune epitopes sites, disordered regions and domain families. 3DBionotes has also been integrated in the output page of other West-Life services such as HADDOCK
Keywords	Bioinformatic Tools
URL	https://3dbionotes.cnb.csic.es/ws
Grid-enabled	No
Cloud-enabled	No
Operational since	March 13th 2017
Total number of users (October 27th 2017)	1094 (unique IPs)
Number of user submissions processed (October 1st 2017)	5411
Number of grid/cloud jobs	Not applicable
Key references	Segura et al (2017) 3DBIONOTES v2.0: a web server for the automatic annotation of macromolecular structures. *Bioinformatics*. btx483

Appendix 5: Portal summary – CCP4-Morda

Portal name	**CCP4-Morda**
Short description	MoRDa is an alternative to two other CCP4online services 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 on-going to provide REST APIs and integration with the Virtual Folder for all CCP4 online portals.
Keywords	Crystallography, molecular replacement; structure database
URL	http://www.ccp4.ac.uk/morda/
Grid-enabled	No
Cloud-enabled	No
Operational since	July 2016
Total number of registered users (October 1st 2017)	372
Number of user submissions processed (October 1st 2017)	2234
Number of grid/cloud jobs	0
Key references	A. Vagin, A. Lebedev, Acta Cryst. (2015). A71, s19

Background information

This deliverable relates to WP5; background information on this WP as originally indicated in the description of work (DOW) is included below.

WP5 Title: Virtual Research Environment
Lead: Alexandre M.J.J. Bonvin (UU)
Participants: STFC, NKI, EMBL, MU, CSIC, CIRMMP, Instruct, UU, Luna, INFN

Work package number	5		**Start date or starting event:**		0
Work package title	Virtual Research Environment				
Activity Type	Support				
Participant number	**1**	**2**	**3**	**4**	**5**
Person-months per participant:	6	3	22	9	27
Participant number	**6**	**7**	**8**	**9**	**10**
Person-months per participant:	24	9	22	22	15

Objectives
This WP is centered on building and operating the VRE web portal that will provide the entry point for users, developers and all other stakeholders. We will build a web portal integrating all already existing and operating services from the various partners and the WeNMR Virtual Research Community (O5.1), and expand it to include new portals, training material and knowledge, and a support center (O5.2, O5.3). In order to better serve the community, customized end-user VMs and/or application containers (e.g. via Docker) will be built for various scenarios (O5.4), to be used on local infrastructures (e.g. within a company) or on the EGI federated cloud resources. Additionally, portals for newly identified applications will be developed and put in production during the project to increase the service portfolio of the VRE (O5.5). The list of objectives is thus:
• **O5.1**: Deployment and operation of the West-Life-VRE portal, integrating all relevant existing services, training and support components (from WeNMR and other partner sites) and extending them.
• **O5.2**: Establishment and operation of the West-Life-VRE support and expertise center for users and software developers, covering all VRE areas. This task will cooperate closely with the relevant EGI-Engage Competence Centers (e.g. MoBrain).
• **O5.3**: Provision of information and training material covering all VRE areas and offered services.
• **O5.4**: Development and integration of new service portals.
• **O5.5**: Provision of customized end-users VMs and/or containers for various applications.

Description of work and role of participants

The above objectives will be addressed through the following tasks:

Task 5.1 – Deployment and operation of the West-Life portal (**Luna**, all)**.**
This task will directly address **O5.1**. It will start by defining the baseline of existing services across all partners (such as X-ray crystallography from CCP4 and the corresponding ones for cryoEM from the CSIC) together with those of the WeNMR VRC. The CSIC will contribute with the Web Services developed at the Instruct Image Processing Center in Madrid, making use of the Web interface of the SCIPION platform for software integration. These will then be integrated into a new VRE portal which will provide end users with a friendly and dynamical entry point to all services, knowledge and support center. The portal will be built on innovative technology developed by LUNA and we aim to migrate when possible existing portals to make direct use of the technology solutions offered by LUNA. In this task, we will also investigate and harmonize user authentication and authorization mechanisms (AAI) (e.g. both the Instruct and the WeNMR sites have user registration mechanisms in place, and WeNMR has implemented a single-sign-on (SSO) mechanism connected to Edugain). The choice and implementation of AAI mechanism will be done in close collaboration with EGI- Engage to maximize compatibility and impact. The new VRE portal will also implement tools and services related to data discovery and access (see WP6).

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

Task 5.3 – Development and integration of new service portals (**UU**, all)**.** This task will directly address **O5.5**. While most of the existing WeNMR portals are already making use of the EGI Grid infrastructure with support from several NGIs within and outside Europe, this VRE project will be adding several portals that are already in place but depend on local and possibly limited resources, as is currently the case for most services for X-ray crystallography and cryoEM. This task will interface those portals (and newly identified ones during the projects) to the most suited e- Infrastructure solution(s), being it grid, CLOUD or HPC resources. Note that we will benefit here from the interaction with various Competence Centers under the new EGI-Engage project, specially the MoBrain Competence Center, to which several partners of West-Life VRE participate (UU, CSIC, CIRMMP and STFC) . Care will be taken to offer user-

friendly interfaces, with a VRE- integrated AAI. The most suited submission mechanisms will be selected. For example, we might adopt the efficient DIRAC4EGI service, but could also build on CLOUD and desktop grid (crowd computing) resources offered by the International Desktop Grid Federation (IDGF). A commercial service will also be offered by LUNA for users (both for profit and non-profit) requesting priority access to resources.

Task 5.4 – **Customized end-users VMs** (**STFC**, all). This task will directly address **O5.5**. Structural biology research has been targeting increasingly larger macromolecular machinery of the cell. Consequently, researchers need access to a wide range of techniques and expertise in order to truly exploit structural biology data. In most cases, however they are expert in only one or a few techniques and associated software. In this task we will build custom VMs for different use cases, with all the necessary software, documentation and examples. Thanks to their suitably designed customization, these VMs will be useful not only to expert structural biologists but also to researchers who want to exploit structural biology as a tool to gain insight in their biological/biomedical research. Different VM types and/or application containers (e.g. via Docker) will be provided, to allow use on both the EGI Federated Cloud and OpenStack/Nebula resources for example, but also local installation on a user's laptop (e.g. with VirtualBox and VMware). This will also potentially be an attractive mechanism for offering commercial services to companies, on their own internal infrastructure when IP issues are preventing external use.

Deliverables

No.	Name	Due month
5.1	Project portal	3
5.2	Overview (baseline) of services and portals to be integrated into the new VRE	4
5.3	Prototype of the new VRE portal functionality	6
5.4	Report on activities of the Helpdesk	18
5.5	VRE-integrated PDBe search and query API's	18
5.6	Report on available VMs with associated documentation/use case for each of them	24
5.7	Report on access and usage statistics of the various services	24
5.8	Report on access and usage statistics of the various services	36
5.9	Update Report on activities of the Helpdesk	36