# Deliverable D5.8

| | | |
|---|---|---|
| Project Title: | World-wide E-infrastructure for structural biology | |
| Project Acronym: | West-Life | |
| Grant agreement no.: | **675858** | |
| | | |
| Deliverable title: | Final 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: | 36 | |
| Actual delivery date: | 36 | |
| 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:** | B. Jimenez Garcia / M. Trellet / A. Bonvin | UU/WP5 | 27.10.2018 |
| **Approved by:** | | | |

## DOCUMENT LOG

| ***Issue*** | ***Date*** | ***Comment*** | ***Author/Partner*** |
|---|---|---|---|
| **v.1** | 23.10.2017 | First version | B. Jimenez Garcia / M. Trellet / A. Bonvin / UU |
| **v.2** | | Second version with input from all partners | B. Jimenez Garcia / M. Trellet / 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 started from a leading position in the Structural Biology field, in that its partners were already providing valuable 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 providing usage statistics for 2017 in Deliverable 5.7, we are now providing statistics for 2018 for all services in operation at this date. Furthermore, we are providing statistics for three new services: CCP4-SIMBAD, DipCheck and PRODIGY-Ligand.

— Comparing the statistics and their associated KPIs over the three years of the project, we can observe a sustained growth in both users and usage for the majority of services, which translates into a sustained overall 25-40% increase rate per year.


# 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

## 1.1 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, 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. By now, West-Life has been operating for almost three 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. One of those is a newly developed portal added to the West-Life portfolio.

## 1.2 Description of new portals

The pre-existing services provided by the West-Life partners have been described in Deliverable 5.7. Here we briefly summarize the new services introduced since the publication of Deliverable 5.7 a year ago. These consist of three new services: CCP4-SIMBAD, which is a pipeline for solving crystal structures by sequence-independent molecular replacement, DipCheck, which is a structure validation tool; and PRODIGY-Ligand, which is an extension of the PRODIGY web portal to support binding affinity prediction for protein-small molecule complexes.

### 1.2.1 CCP4-SIMBAD

SIMBAD is a pipeline for sequence-independent molecular replacement, good for identifying if a crystal contains a contaminant protein. It can also help to find homologous structures in difficult-to-solve novel target cases (see Appendix 1). It implements a 3-stage protocol consisting of (1) a lattice parameter search of the PDB, (2) molecular replacement using a small set of common crystallisation contaminants, and (3) molecular replacement using the MoRDa database of representative domain structures. The portal was in test phase the previous period but now officially in production.

### 1.2.2 DipCheck

DipCheck is a validation tool for protein backbone geometry, developed by Joana Pereira and Victor Lamzin at EMBL Hamburg. The tool uses a Euclidean 3D space (DipSpace) of the orthogonal descriptors of the geometry of a 5-atom dipeptide unit: CA(i-1)-O(i-1)-CA(i)-O(i)-CA(i+1). The DipSpace database contains 1,024,000 data points derived from the selected set of 1,300,000 dipeptide fragments from the well-refined structures deposited in the PDB.
DipCheck classifies the geometry of the middle, CA(i) atom in four categories: Favoured region, Allowed region, Generously allowed region and Disallowed region. DipCheck also classifies the overall geometry of a protein model, according to its DipScore distribution, in the same four categories. DipCheck thus adds a structure validation tool to the WestLife services portofolio (see Appendix 2).

### 1.2.3 PRODIGY-Ligand

PRODIGY-Ligand is an extension of our PRODIGY web-server for the prediction of binding affinity in protein-protein complexes. By using a combination of simple structural properties, such as the residue-contacts made at the interface, PRODIGY has demonstrated a top performance compared to other state-of-the-art predictors in the literature.
PRODIGY-Ligand is an extension of it, aimed at the prediction of affinity in protein-small ligand complexes. The predictive method, properly readapted for small ligand by making use of atomic instead of residue contacts, has been successfully applied for the blind prediction of 102 protein-ligand complexes during the D3R Grand Challenge 2. PRODIGY-LIG has the advantage of being easy to use, generic and applicable to any kind of protein-ligand complex. It provides an automatic, fast and user-friendly tool ensuring broad accessibility. For further details see Appendix 3.

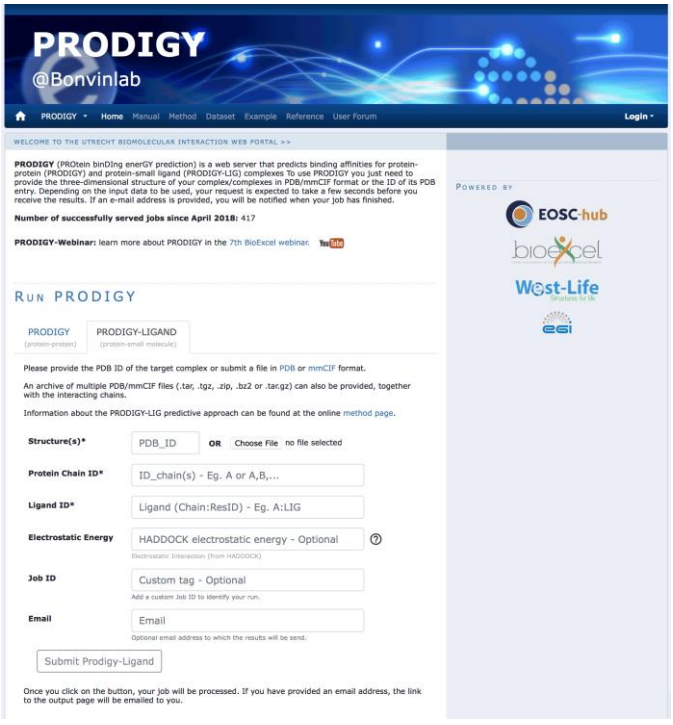


**Figure 1:** View of the new PRODIGY-Ligand web portal.

## 1.3 Portal Statistics

The following table reports on the total statistics of all portals aggregated over the entire duration of the project.

| | Users total | New users | Job submissions | Grid/cloud jobs | %increase #Users total | %increase #new users | %increase #Job submissions | %increase #Grid/cloud jobs |
|---|---|---|---|---|---|---|---|---|
| Baseline 2015 | 16788 | 4437 | 43090 | 7692083 | n.a. | n.a. | n.a. | n.a. |
| TOTAL 2016 | 21218 | 5534 | 52914 | 9246639 | 26% | 25% | 23% | 20% |
| TOTAL 2017 | 24985 | 8081 | 72684 | 8935684 | 18% | 46% | 37% | -3% |
| TOTAL 2018 | 35695 | 10566 | 85800 | 7501906 | 43% | 31% | 18% | -16% |

The percentage reported are always with respect to the previous period.

**Overall we can observe an increasing trend in both number of new users using the WestLife portals per year and number of job submitted indicating a sustained increased usage of our services over all years.**

The only decreased observed is in the number of grid/cloud jobs submitted. This can be explained by some shifts in portal usage. However, while the total number of jobs submitted to HTC resources has slightly decreased in 2018, it is still at a high level of 7.5 million jobs per year. Note also that actually the total CPU usage on the grid has increased in 2018 from around 18 to 21 million of normalized CPU hours, as reported in D4.6. So while the total number of HTC jobs has decreased in 2018, their average duration has increased.

Below, we report usage statistics of the West-Life associated services including, where applicable, the number of users and job submissions for 2016, 2017 (both reported in previous deliverables) and for 2018, with statistics extrapolated to the full 2018 (the statistics were collected until late September, beginning of October and were extrapolated to the entire 2018 year).

Two services have been discontinued in 2018: UNIO, because of a lack of user interest, and GROMACS for which a cloud version has been developed by the MU partner but not yet put into production. Those were already flagged in the 2017 statistics (in red) and are no longer shown in the 2018 statistics.

KPIs are presented for each period compared to the previous year. New services for a given year 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.

### 1.3.1 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 | 26% | 24% | 22% | 20% |

### 1.3.2 Portal statistics 2017

| | Stats 2017 | | | | KPIs 2017 | | | |
|---|---|---|---|---|---|---|---|---|
| | Users total | Users 2017 | Job submissions | Grid/cloud jobs | %increase #Users total | %increase #Users 2017 | %increase #Job submissions | %increase #Grid/cloud jobs |
| AMPSNMR | 468 | 78 | 454 | 1383 | 20% | -13% | 145% | -62% |
| ARPWARP | 4974 | 894 | 5362 | 0 | 13% | 34% | 55% | n.a. |
| AUTORICK | 2893 | 461 | 4517 | 0 | 13% | 7% | 4% | n.a. |
| CCD | n.a. | n.a. | 4412 | 0 | n.a. | n.a. | 10% | n.a. |
| CCP4-AMPLE | 237 | 205 | 708 | 0 | 27% | 10% | 32% | n.a. |
| CCP4-BALBES | 1581 | 823 | 3800 | 0 | 4% | 18% | 21% | n.a. |
| CCP4-Crank2 | 123 | 112 | 601 | 0 | 31% | 19% | 2% | n.a. |
| CCP4-MORDA | 475 | 457 | 1775 | 0 | 135% | 126% | 125% | n.a. |
| CCP4-MRBUMP | 918 | 509 | 1837 | 0 | 5% | 12% | 7% | n.a. |
| CCP4-SHELX | 164 | 134 | 645 | 0 | 9% | -11% | -2% | n.a. |
| CCP4-SIMBAD | 72 | 70 | 143 | 0 | 380% | 367% | -3% | n.a. |
| CCP4-ZANUDA | 410 | 222 | 554 | 0 | 5% | 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. |
| PRODIGY-Ligand | | | | | | | | |
| POWERFIT | 99 | 78 | 150 | 96 | 371% | 271% | 341% | 638% |
| SpotOn | 105 | 105 | 165 | 0 | n.a. | n.a. | n.a. | n.a. |
| SCIPION | n.a. | n.a. | 443 | 40170 | n.a. | n.a. | 6% | 32% |
| 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. |
| 3DBIONOTES | 1406 | 1406 | 6954 | n.a. | n.a. | n.a. | n.a. | 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 | 24985 | 8081 | 72684 | 8935684 | 18% | 46% | 37% | -3% |

### 1.3.3 Portal statistics 2018 (extrapolated)

| | Stats 2018 (extrapolated) | | | | KPIs 2018 | | | |
|---|---|---|---|---|---|---|---|---|
| | Users total | Users 2018 | Job submissions | Grid/cloud jobs | %increase #Users total | %increase #Users 2017 | %increase #Job submissions | %increase #Grid/cloud jobs |
| AMPSNMR | 578 | 110 | 2482 | 13277 | 23% | 40% | 447% | 860% |
| ARPWARP | 5789 | 815 | 8492 | 0 | 16% | -9% | 58% | n.a. |
| AUTORICK | 3247 | 354 | 3765 | 0 | 12% | -23% | -17% | n.a. |
| CCD | n.a. | n.a. | 8850 | 0 | n.a. | n.a. | 101% | n.a. |
| CCP4-AMPLE | 342 | 226 | 764 | 0 | 44% | 10% | 8% | n.a. |
| CCP4-BALBES | 1955 | 774 | 3213 | 0 | 24% | -6% | -15% | n.a. |
| CCP4-Crank2 | 159 | 107 | 442 | 0 | 29% | -4% | -26% | n.a. |
| CCP4-MORDA | 557 | 499 | 1783 | 0 | 17% | 9% | 0% | n.a. |
| CCP4-MRBUMP | 1165 | 494 | 1401 | 0 | 27% | -3% | -24% | n.a. |
| CCP4-SHELX | 231 | 138 | 365 | 0 | 41% | 3% | -43% | n.a. |
| CCP4-SIMBAD | 102 | 57 | 169 | 0 | 42% | -19% | 18% | n.a. |
| CCP4-ZANUDA | 535 | 212 | 472 | 0 | 30% | -4% | -15% | n.a. |
| DISVIS | 606 | 446 | 646 | 537 | 279% | 269% | 17% | 146% |
| FANTEN | n.a. | n.a. | 821 | 0 | n.a. | n.a. | -34% | n.a. |
| HADDOCK | 12199 | 2009 | 28081 | 7396787 | 20% | 7% | -16% | -17% |
| PDB-REDO | 1800 | 709 | 5763 | 0 | n.a. | 52% | 16% | n.a. |
| PRODIGY-Ligand | n.a. | n.a. | 577 | 0 | n.a. | n.a. | n.a. | n.a. |
| POWERFIT | 416 | 317 | 236 | 168 | 321% | 307% | 57% | 75% |
| SpotOn | 456 | 351 | 8047 | 0 | 334% | 234% | 4777% | n.a. |
| SCIPION | n.a. | n.a. | 649 | 21492 | n.a. | n.a. | 46% | -46% |
| VICI | 695 | 56 | 52 | 0 | 44% | -18% | -47% | n.a. |
| 3DBIONOTES | 4418 | 2675 | 8704 | n.a. | 214% | 90% | 25% | 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 | 35695 | 10566 | 85800 | 7501906 | 43% | 31% | 18% | -16% |

# Appendix 1: Portal summary – CCP4-SIMBAD

| | |
|---|---|
| **Portal name** | **CCP4 online - SIMBAD** |
| **Short description** | SIMBAD is a pipeline for sequence-independent molecular replacement, good for identifying if your crystal contains a contaminant protein. Can also help to find homologous structures in difficult-to-solve novel target cases |
| **Keywords** | Crystallography, molecular replacement |
| **URL** | http://www.ccp4.ac.uk/ccp4online |
| **Grid-enabled** | No |
| **Cloud-enabled** | No |
| **Total number of registered users** | 102 (since August 2016) |
| **Key references** | Simpkin, A. J., Simkovic, F., Thomas, J. M. H., Savko, M., Lebedev, A., Uski, V., Ballard, C., Wojdye, M., Wu, R., Sanishvili, R., Xu, Y., Lisa, M.-N., Buschiazzo, A., Shepard, W., Rigden, D. J. & Keegan, R. M. (2018). SIMBAD: a sequence-independent molecular-replacement pipeline. Acta Crystallogr D Struct Biol., 74, 595-605. |

# Appendix 2: Portal summary – DipCheck

| | |
|---|---|
| **Portal name** | **DipCheck Validation** |
| **Short description** | DipCheck is a validation tool for protein backbone geometry, developed by Joana Pereira and Victor Lamzin, EMBL Hamburg. The tool uses a Euclidean 3D space (DipSpace) of the orthogonal descriptors of the geometry of a 5-atom dipeptide unit: CA(i-1)-O(i-1)-CA(i)-O(i)-CA(i+1). The DipSpace database contains 1,024,000 data points derived from the selected set of 1,300,000 dipeptide fragments from the well-refined structures deposited in the PDB.<br><br>DipCheck classifies the geometry of the middle, CA(i) atom in four categories: Favoured region, Allowed region, Generously allowed region and Disallowed region. DipCheck also classifies the overall geometry of a protein model, according to its DipScore distribution, in the same four categories. |
| **Keywords** | Integrative modelling; biomolecular structures; validation |
| **URL** | https://arpwarp.embl-hamburg.de/ |
| **Grid-enabled** | no |
| **Cloud-enabled** | no |
| **Operational since** | 1 September 2017. Integrated to ARP/wARP web service on 10 April 2018 |
| **Total number of registered users** | Not available as the user registration was introduced in April 2018 when the DipCheck web service was integrated into ARP/wARP web service |
| **Number of user submissions processed (1 September 2017 to 30 September 2018)** | 158 submissions |
| **Number of grid/cloud jobs** | 0 |
| **Key references** | Pereira J, Lamzin VS. A distance geometry-based description and validation of protein main-chain conformation. IUCrJ (2017) 4, 657-670. |

# Appendix 3: Portal summary – PRODIGY-Ligand

| | |
|---|---|
| **Portal name** | **PRODIGY-Ligand** |
| **Short description** | PRODIGY-Ligand is an extension of our PRODIGY web-server for the prediction of binding affinity in protein-protein complexes. By using a combination of simple structural properties, such as the residue-contacts made at the interface, PRODIGY has demonstrated a top performance compared to other state-of-the-art predictors in the literature.<br>PRODIGY-Ligand is an extension of it, aimed at the prediction of affinity in protein-small ligand complexes. The predictive method, properly readapted for small ligand by making use of atomic instead of residue contacts, has been successfully applied for the blind prediction of 102 protein-ligand complexes during the D3R Grand Challenge 2. PRODIGY-LIG has the advantage of being simple, generic and applicable to any kind of protein-ligand complex. It provides an automatic, fast and user-friendly tool ensuring broad accessibility. |
| **Keywords** | Drug design; biomolecular structures; protein-ligand complexes; binding affinity |
| **URL** | https://wenmr.science.uu.nl/prodigy/ |
| **Grid-enabled** | no |
| **Cloud-enabled** | no |
| **Operational since** | 1 April 2018. Added as a separate tab on the PRODIGY server |
| **Total number of registered users** | The service does not require registration |
| **Number of user submissions processed (1 April 2018 to 9 October 2018)** | 417 |
| **Number of grid/cloud jobs** | N.A. |
| **Key references** | A. Vangone*, J. Schaarschmidt, P. Koukos, C. Geng, N. Citro, M.E. Trellet, L.C. Xue and A.M.J.J. Bonvin. Large-scale prediction of binding affinity in protein-small ligand complexes: the PRODIGY-LIG web server. Bioinformatics, Advanced Online Publication (2018). |

# 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 |