Deliverable D1.4

Project Title:	World-wide E-infrastructure for structural biology	
Project Acronym:	West-Life	
Grant agreement no.:	**675858**	
Deliverable title:	Draft sustainability work plan	
WP No.	1	
Lead Beneficiary:	1: STFC	
WP Title	Management	
Contractual delivery date:	30-Apr-2017	
Actual delivery date:		
WP leader:	STFC	1
Contributing partners:	Instruct - 7	

Deliverable written by Susan Daenke

Contents

1 Executive summary

In section 3.1, the objectives for achieving sustainability are outlined and section 3.3 provides planned activities and information to be gathered as a foundation on which the consortium will make key decisions that will enable delivery and sustainability of the e-infrastructure over a period of 10 years or more. The sustainability plan 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.

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**		No
2	**Provide automated pipelines to handle multi-technique datasets in an integrative manner**		No
3	**Provide integrated data management for single and multi-technique projects, based on existing e-infrastructure**		No
4	**Foster best practices, collaboration and training of end users**	Yes	

3 Detailed report on the deliverable

3.1 Introduction

A recent report by the European Commission DG-Research and Innovation on the long-term sustainability of research infrastructures (presented at the ESFRI Roadmap 2016 Launch event in Amsterdam), draws on and extends information gathered from several RI sources including EIROforum (members are CERN, EMBL, ESA, ESO, ESRF, EUROfusion, European XFEL and ILL), ESFRI, ERA stakeholders, e-IRG groups and other international and national infrastructure bodies. Data previously collected by EIROforum identified five main criteria that enable RIs to be sustainable[1]**:**

1. Relevance to its **scientific community** and the ability to generate scientific excellence;

2. Sustainable **governance model and legal framework**;
3. Sustainable **funding model**;
4. Ability to attract scientific talent and build a critical mass of scientific expertise;
5. Socio-economic impact.

The 2016 report extended this list to evaluate four additional conditions for sustainability, namely:

6. Unlocking the **innovation** potential of the RI;
7. Exploiting the data generated by the RI;
8. Structuring the international dimension of the RI;
9. Planning and managing the upgrading and decommissioning of the RI.

A sustainability plan for West-Life needs to address a long-term delivery system for the services offered to the user communities. This may include West-Life as a standalone service for the delivery of tools and resources, or its integration with one or more research infrastructures or other e-infrastructures that expand and sustain the service for the long term. Intrinsic to this aim are solutions that provide ease of use (for example via a single portal for access), seamless access to related infrastructure services (such as common SSO implementation), and a strategic governance plan that has the ability to shape the future development and implementation of West-Life in response to scientific and technical demand.
In this draft plan, we have used these criteria to identify the four priority areas where strategic and practical solutions will form the basis for sustainability; specifically, we have reformulated criteria 1, 2, 3 and 6 above for the case of West-Life VRE. The four areas are: the user community; the service provision; the governance and funding model; innovation. These are described more fully in section 3.2 of this document.
The criteria 4, 5, 7, 8 and 9 relate more directly to infrastructures that have taken the key initial decisions (see above) towards a sustainability plan and can build on the options that become available as a result of these early decisions. The West-Life final sustainability report (D1.5) will review these criteria within the framework of the plan.

3.2 Description of the work package task and the deliverable

3.2.1 Objectives

West-Life provides a portal for digital services that support structural biology researchers. Services include computing, data storage and visualization systems that enable integration of multiple approaches and experimental techniques in the structural biology field (including X-ray crystallography, cryo-EM, NMR, SAXS). An integrated technical approach is pivotal for understanding complex intracellular interactions at high resolution, in real time and in the cellular context. Techniques complementary to those provided through West-Life are available through a number of other infrastructure resources: Instruct provides a full catalogue of technical infrastructure that can provide a workflow through multiple structural techniques; data deposition is supported by the Protein Data Bank and associated federated databases such as EMDB, BMRB and SASBDB depending on the structure determination techniques used; iNEXT

provides access to structural techniques for projects that specifically advance translational research.

West-Life aims to provide the means to store and process integrated datasets derived from different techniques in order to build robust structural models of molecular machinery in systems relevant to human health and disease. West-Life brings together existing data management facilities and is initiating new processes that expand the opportunities available for processing data, including capture and preservation of metadata.

Structural biology technologies advance at a fast pace and the related data management processes should advance in sync. The structural biology community needs a robust digital infrastructure to enable the use and reuse of data obtained via multiple routes. The West-Life infrastructure should provide continuity of service underpinned by reliability, excellent user support and quality standards of e-infrastructure services that improve the quality of research outcomes. Together, these features form the basis of sustainability for an infrastructure.

Deliverable 1.4: This draft work plan elaborates the information and processes that should be established in order to achieve sustainability for West-Life.
The full sustainability report will form Deliverable D1.5, to be submitted in month 36.

3.2.2 European policy drivers for e-infrastructures

The Juncker plan for structural funds and the H2020 work programme identified e-infrastructures as vehicles to deliver 'open, data-intensive and networked research as a driver for faster and wider innovation', underlining that e-infrastructure is one of the key elements for research and innovation focused on data or benefiting from it.
E-infrastructures should ensure scientific excellence and widen access to important research and education resources. They should be:

- User centric
- Service oriented
- Innovative
- Sustainable

E-infrastructure platforms and services need to evolve through innovation to respond to the long-term needs of research communities, industry and public communities at large. Innovation can be platform-driven and/or user-driven and can be achieved partly by better co-operation and integration of e-infrastructure services. Integration helps e-infrastructure services to be stable and funded with no temporal gaps and supports persistent and reliable operation. Further, e-infrastructure service and development should be in line with the European policy context of Open Science, including open access to data.
Based on these drivers, the West-Life sustainability plan will establish a model that defines the following four components of a sustainable model:

1. The user communities and the demand for services

2. The main service provision and the process thereof

3. The governance and funding model that will address the medium term and long term e-infrastructure sustainability

4. Strategies for innovation and integration with the service

3.2.3 Measures for achieving sustainability – the draft outline plan

In the remaining period of the project, information will be gathered to populate the sustainability plan with definitions and examples of how West-Life will fulfill the criteria for sustainability identified above. Many of the listed activities are happening now as part of the current West-Life project, but are also important for sustainability of West-Life beyond the current funding.

Area	Objectives	Activities to support sustainability
1. User community and the demand for services	To identify user communities	a. Exploit methods to attract existing and new users: Easy access web interfaces (identifiably branded); exploit public resources: e.g. YouTube for publishing workshop lectures and videos/movies; Twitter (@WestlifeSB) and LinkedIn for e-infrastructure partners and users (D2.3 Report on community engagement with the structural biology community; D2.4 Report on the industrial user engagement; D2.5 Summary engagement report); service the computational structural biology community (prediction of complexes and related infrastructure) alongside the structural biology research community (Task 6.4); b. Gather information on the use of resources by non-structural biology communities (D3.2, D3.3, D3.5, D3.6 all address the demand for data services from other life sciences research infrastructures in the ESFRI group); c. collect information through user registration and gain access to mailing lists through related and joint RI activities; compile metrics on website hits, data/tools jobs and Twitter followers; d. Undertake surveys to establish user profiles; compile a user community scale and profile from publication citations; e. Use project or external outreach programmes to provide information on West-Life services and gather information on potential users from these activities.
	To identify user demands of services	a. Collect feedback from users of the e-infrastructure; b. Undertake periodic peer review of services and obtain advice on new service provision that will add value; c. Host training activities to support the existing services and introduce new tools and resources;

		d. Undertake surveys to gather information on services that could be expanded; into new communities e.g. industry, non-structural biology communities, meta-analysis project groups; needs for data storage and how data is being used; e. Sustain a support network for users that meets demand and optimizes research outcomes: e.g. help centre, forums, chat, blog, Wiki, Trac, wizards, guided tours.
2. Main service provision	To provide integration of existing services	a. Exploit links with EUDAT - EUDAT has broad transversal reach into many communities across life sciences, physical sciences, social and environmental sciences (https://www.eudat.eu/synergies); b. Develop links with EIROforum (http://www.eiroforum.org/) which has reach into Europe's intergovernmental research organisations operating large scale infrastructures in physical and life sciences (e.g. for life sciences: EMBL, ESRF (http://www.esrf.eu/), XFEL, ILL). EIROforum identifies and quantifies IT infrastructure (such as high-speed international and global networks) which is a critical and common foundation of research activity. c. Integrate services with the European Open Science Cloud (EOSC) – federation of existing infrastructures (ESFRIs) and scientific clouds – through EUDAT; d. Link with European Data Infrastructure – developing and deploying large scale HPC, data and network infrastructure; e. Work with and build in solutions provided by BioExcel (http://bioexcel.eu/), INDIGO-DataCloud (https://www.indigo-datacloud.eu/), MoBRAIN (https://mobrain.egi.eu/), Phenomenal VRE (http://phenomenal-h2020.eu/), EGI-engage (https://wiki.egi.eu/wiki/EGI-Engage), OneData (https://onedata.org) , B2SHARE (https://b2share.eudat.eu/); f. Integrate with services provided e.g. through OpenAIRE which handles Open Access for publications, research data, data sources etc (https://www.openaire.eu/); g. Establish quality metrics to characterise individual services (the reliability and availability of a service is essential for judging the extent to which it can be depended upon) and implement a monitoring plan which collects and reports (if not derives) quality metrics to the user communities.

		These activities will include work undertaken in WP2 and WP3.
	To provide and maintain an efficient core process to deliver services to user communities	a. Establish the core service provision (i.e. the VRE process) as the basis of the sustainable infrastructure. Ensure this is stable, well supported, user friendly, regularly updated, scalable, compatible with linked resources (WP4 activities); b. Provide a clear and simple path for new structural biology services to be integrated into the infrastructure c. Adopt and implement a federated ID management system to allow users to access tools and services from all service providers; support and enable a single access point for users which integrates with complementary infrastructures and other e-infrastructures or tool/data repositories (D4.2); d. Evaluate commercial services to augment the VRE: e.g. major commercial cloud providers, Dropbox, OneDrive; e. Lobby for hardware resources on behalf of the structural biology community, in particular explaining the computational needs that must be met to gain full value from new instruments including electron microscopes. f. Establish service standards that define efficiency and excellence; monitor this through Exchange of Experience workshops with related e-Infrastructures and complementary research infrastructures; g. Provide use cases as examples of service delivery workflows; h. Monitor standards through user feedback, publications, peer review.
3. Governance and Funding model for sustainability	To establish a legal framework with effective governance to sustain the e-infrastructure for 10 years or more	a. Establish a governance model for service provision which is inclusive for providers and users; ensure the governance bodies have representation from members/funders, users, service providers, the legal entity (e.g. Board of Directors, if required). Bodies should: provide strategic direction; operational oversight; legal and financial oversight. Subcommittees can be convened. Establish rules of procedure for major governing bodies. b. Evaluate and identify a legal model that is most appropriate for the e-infrastructure: e.g. ERIC, EEIG, organisation (limited company, GmbH, non-profit organization e.g. ASBL/VZW). For sustainability, this model should allow direct employment of staff, financial/VAT registration and scrutiny, appropriate timeframe (e.g. 10 years or more), ability to enter into contracts (including E=with the EC);

		c. Confirm the branding of West-Life as a separate entity, or as a component of wider initiatives (e.g. Instruct) d. Identify an operational model for the VRE – virtual, distributed service providers with a Hub administrative office; e. Build in the roles of other e-infrastructures such as EUDAT, EGI, GEANT and PRACE which will provide longevity in securing the processes for the interchange of data between e-infrastructures and research infrastructures and common users; enable representation on governing bodies for transfer of information. f. Establish service level agreements with providers of services and other agreements for protection or disclosure of data, intellectual property, licensing;
	To establish a financial model that provides sustainable funding for the core service process with opportunities for further funding streams for expansion/innovation.	a. From the legal model, establish the route of core funding to underpin the core e-infrastructure services (the process). Obtain funding commitments from funding bodies/members for 5 years to allow financial forecasting and risk management; b. Establish the priorities of funding derived from public funds, membership contributions, private/project funds (grants etc), service fees; c. Engage funders to sponsor and promote the activities of the e-infrastructure and use of its services by interested communities and support implementation of new services; d. Identify other funding routes that will allow development of the e-infrastructure e.g. innovative development of technologies or service delivery, training (staff, users, providers) e.g. commercial partnerships, collaborative enterprise, H2020 Framework Partnership Agreements (FGAs); e. Identify services (e.g. cloud services) that can be provided free at the point of service (public or privately funded but made available through a sponsoring organisation);
4. Innovation	To realize the innovation potential of the e-Infrastructure:	a. Identify and prioritise new services or resources that are in demand and plan for their development/implementation, including funding for them (D7.1-7.9); b. Establish/maintain links with industry with meetings/working groups etc to improve integration and the development of commercial applications and tools;

		c. Implement an IPR policy that supports the exploitation of opportunities as they arise and best serves the user communities; d. Support interoperability that drives innovation by enabling simple access to multiple resources, including spanning across scientific boundaries; e. Provide an environment that attracts talented developers and implementers who will also be the innovators within the e-infrastructure.

3.2.4 Technical implementations that address sustainability

In planning the delivery of West-life services, we have installed standardized interfaces and backend/frontend technologies to futureproof the portal.

Technical considerations in Work Package 4

For backend convergence (computing job dispatching), DIRAC4EGI was chosen to replace gLite-based submission for longevity since it has a strong user base and has EGI backing. It also addresses the problem of future decommissioning of Workload Management Systems (WMS). The WP4 perimeter was defined around larger size grid/cloud job dispatching, while dispatch of pre-processing jobs has been left to individual services and portals.

Technical considerations in Work Package 5

For support and operations of the West-Life VRE, a common web interface platform was considered for all West-Life services for use as a primary web UI for users.
However, three problems were identified:

1. West-Life services can have vastly different interface requirements (eg: Scipion Web Tools compared to WeNMR Haddock);
2. West-Life services build on pre-existing services (e.g. from WeNMR), alongside new services that are being developed during the project. Porting all of those to a new common interface platform outstrips the man-power available within this project;
3. West-Life services often have multiple sources of funding, and e-science portal operators hire web developers according to their needs of which West-life services form a small part;

Because of these, a decision was taken to bring together existing ‘disconnected’ services under one umbrella, providing a common “data” infrastructure to improve interoperability and thus make it possible to do new “science” by combining different services in workflows.

The West-Life VRE was therefore defined as a “common roof” that brings together all West-Life individual e-science portals, with a common graphical theme, a common SSO mechanism based on Instruct's ARIA and connected to the SSO solutions developed by EGI Engage, and a common virtual folder mechanism (aggregated data access for users), defined in WP6.
ARIA SSO mechanism was chosen because of the stability conferred by Instruct and their commitment to maintain it in the long term. In addition, django and Angular were used because of their popularity as web frameworks, so that others can contribute or lead the single point of entry web portal development in the future as required.

Technical considerations in Work Package 6

The virtual folder that allows West-Life users to mount their data (currently scattered in Dropbox, Amazon, B2DROP) as one single unified file system, has been developed as a simple mechanism that will only require updates, as the data providers update their API. West-Life will monitor trends in storage solutions at the national and international level (e.g. in the Netherlands SURFDrive services the academic research community, offering generous storage solutions). Other solutions like the ones developed under the INDIGO-Datacloud project will also be

considered if found suitable (e.g. OneData). This flexibility will enable Westlife virtual folder to evolve with trends in storage solutions, as all storage providers are abstracted away.
In summary, the West-Life perimeter has been designed with sustainability in mind and should support the portal requirements at least into the medium term.

3. 3 Conclusions and outlook

The work towards developing the sustainability plan will build on the information provided in the project workplan. The following process will progressively populate and develop the plan as the project progresses:

1. Work delivered by each of Work packages 2 to 7 will be viewed in the context of its relevance to the plan (continuous);
2. Metrics on the use of West-Life services will be gathered to set a baseline for demand and to guide strategies for expanding their use (M30);
3. WebEx meetings between STFC/Instruct and WP participants will be held regularly to discuss incorporation of new information into the plan as it is delivered (monthly);
4. A draft plan will be broadly circulated in Year 3 in order to capture all relevant information from participants and a final draft will be submitted to the Executive Steering Board for approval in M30. The final D1.5 report will include the approved Sustainability Plan with an Executive Summary (M36).

References cited

1. EIROforum discussion paper: Long-term sustainability of Research Infrastructures: http://eiroforum.org/science_policy/news/2015/03-25-sustainability-of-research-infrastructures/index.html.
2. Strategy Report on research Infrastructures Roadmap 2018, Malaga, Spain. https://ec.europa.eu/research/infrastructures/pdf/esfri/esfri_roadmap/esfri_roadmap_2016_full.pdf.
3. Towards a shared vision of sustainability for research and e-infrastructures. https://www.linkedincom/pub/eudat-project/51/a45/30a
4. Report on the consultation on long term sustainability of research infrastructures, May 2016. Directorate-General for Research and Innovation. https://ec.europa.eu/research/infrastructures/pdf/lts_report_062016_final.pdf.
5. E-infrastructure for the 21st century – one year later. EIROforum IT working Group, 2nd December 2014. www.eiroforum.org/science.../12...e-infrastructures-for-the-21st-century/index.html.
6. Overview on European e-infrastructure. Augusto Burgueno, DG-Connect. Geant General Assembly, 2015. https://ec.europa.eu/research/openscience/pdf/eosc-workshop.../e-infrastructures.pdf
7. Workshop on ‘Future perspectives for research infrastructures advanced communities. Directorate-General for Research and Innovation, 14th November 2016. https://ec.europa.eu/research/infrastructures/pdf/events/future_perspectives_for_ri_report.pdf.