EXCELERATE Deliverable D4.3

Project Title:	ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences
Project Acronym:	ELIXIR-EXCELERATE
Grant agreement no.:	676559
	H2020-INFRADEV-2014-2015/H2020-INFRADEV-1-2015-1
Deliverable title:	The ELIXIR Compute Platform: A Technical Services Roadmap for supporting Life Science Research in Europe
WP No.	WP4
Lead Beneficiary:	20 - CSC
WP Title	Technical Services
Contractual delivery date:	30 September 2018
Actual delivery date:	09 September 2018
WP leader:	Tommi Nyrönen (FI) and Ludek Matyska (CZ)	35 - MU & 20 - CSC
Partner(s) contributing to this deliverable:	EMBL-EBI, CSC, MU, NBIS, CESNET, ELIXIR Hub

Authors and Contributors:

Steven Newhouse (EMBL-EBI), Tommi Nyrönen (CSC - FI), Ludek Matyska (MU - CZ), Mikael Borg (NBIS - SE), Jinny Chien (EMBL-EBI), Jarno Laitinen (CSC - FI), Ilkka Lappalainen (CSC - FI), Mikael Linden (CSC - FI), Michal Prochazka (MU - CZ), Mirek Ruda (CESNET - CZ), Harri Salminen (CSC - FI), Juha Törnroos (CSC - FI), Christine Staiger (SURFsara - NL), Jonathan Tedds (ELIXIR Hub), Susheel Varma (EMBL-EBI), Tony Wildish (EMBL-EBI)

Reviewers: N/A

Table of contents

1. Executive Summary

This version of the ELIXIR Technical Services Roadmap (ELIXIR-EXCELERATE deliverable D4.3) has been written at the end of PY3 of the ELIXIR-EXCELERATE project (August 2018) and represents an update to the previous version that was established in August 2017 (Deliverable D4.2) and will be updated again in Summer 2019 before the project ends. This version of the roadmap contains a summary of the technical work undertaken during PY3, an assessment of the Platform's capability defined through a number of Technical Use Cases (TUCs), and an assessment of the ability of the ELIXIR Compute Platform (through these TUCs) to support scientific community-driven use cases, and the plans for the future evolution of these integrated services during PY4.

The ELIXIR Compute Platform (ECP) experts have met twice this year to communicate progress in technical service integration driven by the requirements of the ELIXIR scientific communities. The ELIXIR Technical Services Roadmap focuses on requirements of four ELIXIR communities (i.e. marine metagenomics, plants, rare diseases and human data) and training. As these communities also evolve quickly, the roadmap is a living document that provides advice as to the current and future implementation activities, and is subject to change between versions that is publicly accessible and commentable.

The [ELIXIR AAI](https://docs.google.com/document/d/1cJ3mR8lqfZKRMvSFaISmPbqd1OPU-L6YcUFIRnh1rhQ/edit)[1] has grown to be a full production Infrastructure Service with 67 production services and 65 services in the testing environment. ELIXIR AAI is stable and provides a way to federate and integrate scientific services to provide user access management across organisational and country borders by using different sources of identity and structuring these users into groups to support authorisation decisions by individuals service. Use of the ELIXIR AAI has grown beyond internal ELIXIR services to include public ELIXIR services such as the ELIXIR Germany cloud (http://de.nbi), Helix Nebula Science Cloud (to access commercial cloud providers), EUDAT's B2ACCESS service (to access other EUDAT services) and the EMBL-EBI Unified Submissions Interface. Discussions around the planned Life-Science Identity across the BMS RI community have been consolidated within the EOSC-Life proposal which has now been funded.

Moving files (i.e. data) between sites is a key capability for the ECP. The Reference Data Set Distribution Service that is being developed in collaboration with the EUDAT2020 project was introduced into the ECP during PY3 with the support of an ELIXIR IS. Furthermore, new protocols such as FTS3 and htsget, a genome data specific standard have been integrated to the platform. During PY3, the ELIXIR collaboration with the Global Alliance for Genomics in Health (ga4gh.org) has expanded. The ECP deployed GA4GH standards for genomic data transfers in May 2018 and demonstrated transfer sensitive of human datasets to remote secure cloud infrastructure service.

Work within the ECP to integrate the cloud resources affiliated to the ELIXIR Nodes continues. The EGI Federated Cloud model has to date not been widely adopted within the ECP but remains under evaluation by the ECP, especially within the context of the emerging European Open Science Cloud (EOSC) projects EOSC Hub and EOSC-Life, where it may be used as the federation model. The ECP is engaging with EOSC through an ELIXIR Competency Centre that is funded as part of the EOSC-Hub project which builds upon the work that has been undertaken in the EGI-Engage project. ECP leads are also steering the work packages in EOSC-Life for cloud and authentication and authorisation targeted to the community of the Biomedical Science ESFRI.

---

[1]

https://docs.google.com/document/d/1cJ3mR8lqfZKRMvSFaISmPbqd1OPU-L6YcUFIRnh1rhQ/edit

Through work that started in November 2017 and has been continued within an ELIXIR IS in 2018, the ECP has been assessing the use of GA4GH Cloud Workstream standards for providing a cloud based analysis platform. Working the Tools and Interoperability platforms through other IS, the ECP is building a demonstrator that will distribute tasks within a workflow across multiple cloud resources. This work will continue within the EOSC-Life project.

During the PY3, extensive discussions have been initiated to reach a Common ELIXIR resource allocation process and policy to leverage distributed cloud capacities hosted in the ELIXIR nodes. Such a process would allow more straightforward international cloud and data access mechanisms supported by the integrated technologies. Work will continue in this area during the ELIXIR 2019-2023 programme and in the EOSC-Hub and EOSC-Life projects.

Technical discussions with the ELIXIR-EXCELERATE Use Cases (the four scientific use cases and the training activities) continue. Marine Metagenomics, the focus during PY1 continues to develop and consolidate its technical activity. The focus in PY2 on Human Data showcased two key functionalities for secure transfer of sensitive human data and making datasets originating from the European Genome-phenome Archive on a secure cloud. The demonstrator leveraged ECP technologies and GA4GH standards (htsget) for transferring the data, and relied on ELIXIR AAI for user identification and dataset access control. In PY3 the human data community use case continues to be a key scientific driver for technology integration. A new use case relating to plants was added to the focus in January 2018, and by August 2018 the technical experts from WP4 and WP7 had formulated a technology demonstrator focusing on transfer of distributed datasets. The solution delivered through EXCELERATE project effort will be presented in a webinar during the Autumn 2018 based on ECP milestone M4.3.

The basic infrastructure components of the ECP has been defined as Authentication and Authorisation, Data Transfers and Storage, and (hybrid) Cloud services. Development in Container technologies will in 2019-23 add a new task to the ECP agenda. Work will continue to see how these components can be deployed (either directly by ELIXIR Nodes, commercial providers or indirectly in partnership with European e-Infrastructure providers), integrated, optimised and sustained to support European life-science research and to be made available to bioinformatics experts for use to analyse globally significant data resources.

2. Impact

Collaboration within the ELIXIR Compute Platform

- Paper describing Common ELIXIR authentication and authorisation toolkit was published. In the beginning of 2018 ELIXIR AAI enabled 493 institutions whose members can use ELIXIR AAI, integrated 67 production services in total and an additional 65 in testing, and had 1838 users and 370 groups.

- [Public description of the platform roadmap](#)[2] has been distributed extensively, and will be updated to include the technical demonstrators with user communities in PY4
- ELIXIR AAI support has been developed between ELIXIR-CZ, ELIXIR-FI and with helpdesk tool support from ELIXIR-NL. About 500 issue tickets are resolved annually at the moment. Work provides a way to organise ELIXIR AAI support process requests into 1st line support (FAQ), and assign support requests to advanced support teams (2nd line) as necessary. It is possible to apply the same principles to other ELIXIR services.
- ELIXIR Implementation Studies supporting the development of the ELIXIR AAI, deployment and use of storage data transfer services, and the a standards based platform for task based workflows
- Establishing a prototype Reference Data Set Distribution Service (developed with EUDAT2020 support) as part of the ECP

Collaboration within the ELIXIR-EXCELERATE project
- Scientific use cases collaboration has been reported on a separated document
- WP6 marine metagenomics Metapipe service has been deployed and supported outside the original node and made available for scientists
- WP7 technical demonstrator to support plants community has been designed and will be delivered in Q4 2018
- WP9 ELIXIR Compute and Human Data platform demonstrated technology integration across European nodes to transfer sensitive human datasets to remote secure cloud infrastructure service in 2017 and 2018
- A demonstrator with WP8, the Rare Diseases use case, is planned as the final technical integration delivered by the WP4 in ELIXIR-EXCELERATE in 2019.
- WP11 ELIXIR clouds have been made available for bioinformatics training with the additional funding support from an ELIXIR Implementation Study.

Collaboration with European e-Infrastructures projects
- AARC2 project blueprint has been released and ELIXIR AAI endorses and is compatible with the architecture. Life Science AAI requirements document was completed and planned to be deployed with the EOSC. LS AAI is also based on the AARC blueprint.
- ECP participation in the EOSCpilot and EOSC-Hub projects has continued. The EOSC-Life project with the leadership from the ECP in the Cloud and AAI work packages has been funded and will start in March 2019
- Deployment of the marine metagenomics use case (WP6) on cloud resources from both Helix Nebula and the ECP
- A recurring issue for projects and their users is how ELIXIR cloud resources or e-Infrastructure resources can be provisioned. There is no ELIXIR-wide mechanism or policy for handling resource provisioning issues. Typically, each ELIXIR node cloud service operates independently meeting the needs of its local user community using the mechanisms defined by its funding agency.

---

[2] http://drive.google.com/file/d/0B0KXZdVao0kqUE9BbXVrc3ZLY1E/view

3. Project objectives

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

No.	Objective	Yes	No
1	Develop a sustainable and supported research platform for implementing geographically and organisationally distributed Cloud, Compute, Storage and Authentication and Access infrastructure services collected in the ELIXIR registries.	X	
2	Manage external technical dependencies with e-Infrastructures and Nodes with ELIXIR technical coordinator group for services delivered as a priority for the ELIXIR-EXCELERATE Use Cases.	X	
3	Close collaboration with translational, bio-banking and imaging infrastructures at both the European and national level to ascertain that there are effective services to securely access and exchange data	X	

4. Delivery and schedule

The delivery is delayed: •Yes ☑   • No

5. Adjustments made

The deliverable is only slightly delayed in order to produce the final report.

6. Background information

Background information on this WP as originally indicated in the description of action (DoA) is included here for reference.

Work package number	4	Start date or starting event:	month 1
Work package title	**Technical Services**		
Lead	Tommi Nyrönen (FI) and Ludek Matyska (CZ)		

Participant number and person months per participant

1 – EMBL 34.00, 6 - NBIC 2.00 (LTPs: SARA 13.00, RUG 2.00) 8 - CRG 8.5, 12 - BSC 7.00, 17 - INESC-ID 6.00, 20 - CSC 48.00, 23 - UiT 2.00, 25 - SIB 2.00, 26 - CNRS 12.00, 29 - IP 6.00, 35 - MU 40.00, 36 - CESNET 24.00, 38 - DTU 6.00, 41 - ATHENA RIC 8.00, 45 - UU 4.00 (LTP: SU 8.00).

WP4 - Technical Services [Months: 1-48]

MU, EMBL, NBIC, CRG, BSC, INESC-ID, CSC, UiT, SIB, CNRS, IP, CESNET, DTU, ATHENA RIC, UU

The role of the ELIXIR-EXCELERATE Technical Services WP is the practical integration of existing Technical Services available for ELIXIR in the Nodes and e-Infrastructure by testing and contributing to documentation and integration with small-scale programming and scripting where needed. Development is managed outside the WP. As a result of the tasks described below, WP4 will provide a generic integrated platform that can be tailored further for the ELIXIR-EXCELERATE scientific Use Cases (WP6 to 9), Training activities (WP11), and other ELIXIR pilots and projects to meet their specific needs. This includes user support, documentation and guidance to enable and promote technology adoption.

Work Package uses a mechanism of renewal of focus with the ELIXIR Heads of Nodes committee as necessary. If scientific needs change or disruptive technologies emerge that change the technical objectives heads of Nodes committee supports linking of the changed landscape of technical services implementation with the other Work Packages (e.g. ELIXIR resource governance, training, data resources, service registry). Involvement of ELIXIR heads of Nodes is used for securing physical information technology resources from Nodes, and making experts available for collaborative work.

Objectives

The Technical services Work Package (WP) links the ELIXIR scientific programme 2014-2018 to the day-to-day technical service work in the distributed Nodes. The research platform for life science will be achieved through the following objectives:

• Develop a sustainable and supported research platform for implementing geographically and organisationally distributed Cloud, Compute, Storage and Authentication and Access infrastructure services collected in the ELIXIR registries.

• Manage external technical dependencies with e-Infrastructures and Nodes with ELIXIR technical coordinator group for services delivered as a priority for the ELIXIR-EXCELERATE Use Cases.

• Close collaboration with translational, bio-banking and imaging infrastructures at both the European and national level to ascertain that there are effective services to securely access and exchange data.

Work Package Leads: Tommi Nyrönen (FI) and Ludek Matyska (CZ)

Description of work and role of partners

Task 4.1: Leadership (53PM)

Subtask 4.1.1: Management and Coordination (26PM)

This task is responsible for coordinating technical work in the ELIXIR-EXCELERATE project and wider ELIXIR research infrastructure with WP12 building on the emerging community of technical experts in the ELIXIR task forces. In addition, the task establishes appropriate management and technical interfaces into the services and organisations the technical activities are dependent upon.

Partners: FI, CZ

Subtask 4.1.2: Provide a gateway to use European e-Infrastructure services for ELIXIR (GÉANT, EGI, EUDAT, PRACE) (13PM)

Regular requirements gathering from the Use Cases in WP6 to 9 and elsewhere in the ELIXIR community will define biological information service requirements and identify areas and activities that could be sourced by the European e-Infrastructures. Any planned service integration into the ELIXIR Technical Services will be identified in the regular Roadmap documents that will define a technical architecture and technology insertion roadmap. This should include defining the relevant 'account managers' in each public sector e-Infrastructure.

Partners: FI, EMBL-EBI

Subtask 4.1.3: ELIXIR technical community building and knowledge exchange (14PM)

Task grows the community of ELIXIR branded resource providers and sustains community of ELIXIR technical experts (i.e. ELIXIR technical coordinators and Node personnel) through engagement in major e-Infrastructure events, technical workshops, audio/video conferencing and other collaboration mechanisms. Working groups and task forces bring in relevant experts from outside ELIXIR such as e-Infrastructures. This task will be made in collaboration with WP1 (Tools) and WP12 (Management).

Partners: ELIXIR Nodes

Task 4.2: User Facing Support (71PM)

This task interacts with the individuals and projects that are users of the ELIXIR Technical Services platform through their defined Use Cases. The main consumers are the ELIXIR-EXCELERATE Use Cases (WP6 to 9) and other ELIXIR activities(e.g. ELIXIR-EXCELERATE WPs, ELIXIR pilots, and external projectslike EC funded Centres of Excellence or Virtual Research Environments).

Subtask 4.2.1: Technical requirements (28PM)

Gather and analyse the technical requirements for the Technical services platform in order to define the detailed technical specifications and interfaces of the technical service platform. One outcome is the classification of the different Use Cases with technical terms (e.g. small compute, large data input-output; large compute, data access management, etc.).

This work will feed into WP12 concerning requirements requested and procured from any external service providers.

(a) EXCELERATE Use Cases WP6 to 9.

(b) ELIXIR tools registry, ELIXIR training events, data transfers to/from ELIXIR data resources, and authentication and authorization.

Partners: EMBL-EBI, FI, CZ, ES, NL, NO

Subtask 4.2.2: User support and integration (43PM)

Provide a support structure that can be applied to adopters of the ELIXIR Technical Services. This will be focused on the use of ELIXIR Technical platform e.g. for supporting organizing a training event. Task provides operational support for ELIXIR-EXCELERATE activities and externally funded ELIXIR activity (technical pilots and projects). This will take place through a single-point-of-contact 'helpdesk' function and 'hackathons' where users and the providers of the ELIXIR Technical Services work together to integrate functionality across AAI, cloud and data. As a result of this work a set of 'recipes' focused around user activities will be collected into a 'cook book' to enable community adoption (e.g. to run a Galaxy workflow environment on an ELIXIR-affiliated Cloud Resource with accounting if necessary).

Partners: EMBL-EBI, FI, SE, FR, NL

Task 4.3 Technical infrastructure integration (102.5PM)

This task focuses on the integration of Technical Services being delivered by individual ELIXIR Nodes and from the public e-Infrastructures in Europe to meet the requirements of the ELIXIR community (e.g. by establishing account manager relations with each e-Infrastructure). The strategy within this task is to focus on the integration of existing mature and stable services to ensure that these services are easier to uptake by bioinformatics Use Cases (WP6 to 9).

Subtask 4.3.1: ELIXIR AAI - Authentication, authorization (access) integration (28PM)

ELIXIR needs a service based on European federated identity that authenticates an individual is a member of a group (or has a particular role within a group) that can be managed remotely. Group management needs to enable delegate decision making to multiple individuals within a particular community (e.g. institutional representative within a project) and queries from other services.

(a) Establishing an ELIXIR Identity: Federated identity technologies are fairly mature, as are many of the related tools (e.g. REMS, PERUN). This task ensures that ELIXIR research community is fully covered (including users whose home organization does not provide federated identities) and acts as a single IdP for ELIXIR branded services technical work. The task continues to integrate the existing services and ensures that they provide the interfaces needed for adoption within this Work Package, the project and externally (e.g. BMS RIs).

(b) Providing additional AAI services: eduGAIN IdPs, Common IdPs, guest login, Proxy IdP, ELIXIR directory, Attribute self-management for users, Bona fide researcher management, Group/role management, Dataset authorisation management, Credential translation.

Partners: CZ, FI, EMBL-EBI, NL

Subtask 4.3.2: Cloud and Compute integration (42.5PM)

European and national compute centres service clusters and access to resource with open-source cloud technologies is growing. This task integrates the willing services to ELIXIR registry. The way how e.g. IaaS resources are consumed in the research community typically takes place on science-specific platforms and workflows. As a priority WP4 secures resources to support the scientific software workflows for the Use Cases WP6 to 9 and WP11 using the software environment workflows they have chosen for their data analysis framework (e.g. supporting provision of Galaxy as a service for marine metagenomics pipeline).

(a) ELIXIR Cloud accounts: Integrate willing providers (e.g. Embassy Cloud), national level (e.g. CSC, SURFsara, Nordic Secure Cloud, MetaCentrum and CERIT-SC) and regional level (e.g. GÉANT, Helix Nebula and EGI cloud resources) and in the commercial sector (e.g. Amazon, Microsoft, Google). Mechanisms are needed to calculate virtual access costs that can be passed on to projects or funding agencies. Key target is to make accounts to provide resources for WP6 to 9 and WP11 activities.

(b) Enable SME access to ELIXIR cloud resources. We will support billing models such as monthly fee for service subscription or allocation-based costs when free (pre-paid) access is not available. Cost models will be developed with WP12.

Partners: NL, CZ, FR, EMBL-EBI, FI

Subtask 4.3.3: Storage and data transfer (22PM)

Data push and pull is needed in WP6 to 9 supported with commonly agreed technical tools and interfaces. Various transport mechanisms (e.g. GridFTP, http, Aspera, UDPipe, iRods) can be used to move the data to or from Data Resources (WP3). The managed access integration will be piloted in the ELIXIR-EXCELERATE Use Case WP9.

Collaboration with GÉANT (e.g. bandwidth-on services) could be used to provide dedicated network links (e.g. lightpaths) for regular or large data transfer activities between the Nodes. Three common uses will drive WP4 storage and data transfer activities:

(a) Data replication (an updated dataset being moved to multiple remote locations) and data submission (where a dataset is made available for subsequent retrieval and remote analysis). In the former the data source triggers data movement to data sink(s) (e.g. using Globus Transfer) using a replication policy around the data and updates any relevant data catalogues (e.g. B2FIND).

(b) Service to pull relevant datasets for detailed analysis (e.g. Galaxy running on ELIXIR-affiliated cloud resource during training event). The retrieved dataset may be discarded after processing and just the results are retained based on the assumption that the original data will remain accessible for re-analysis.

(c) Data location services will be used to manage and discover data replicas within ELIXIR sites (using technologies such as B2FIND or the EGI Data Catalogue). AAI mechanisms and workflows (e.g. REMS) are needed for gaining approved access entitlements in collaboration with the responsible granting bodies such as data access committees (e.g. EGA).

Partners: SE, EMBL-EBI, ES, CZ, FI

Subtask 4.3.4: Service Registry (10PM)

Integrate with WP1 and WP3 service registry and existing e-Infrastructure registries to enable a wide range of ELIXIR services and resources (e.g. cloud, storage, datasets) so that they become discoverable entities. The service registry provides a 'gateway' by which service providers can advertise and the users consume services. The service registry needs to provide a 'service discovery' function for consumers, but also written advice and requirements on how service providers can advertise their services.

Partners: EMBL-EBI, CZ

7. Appendix 1: The ELIXIR Compute Platform: A Technical Services Roadmap for supporting Life Science Research in Europe

1. Introduction

The ELIXIR Compute Platform (ECP) is being put in place, with the support of the ELIXIR-EXCELERATE project, to support the ELIXIR Scientific Programme for 2014-18 and prepare the forthcoming 2019-23 programme. This deliverable is synchronised with the planned work in the ELIXIR 2019-23 programme, where ECP is recognised as one of the five major ELIXIR platforms. The ECP is a generic platform that provides support to ELIXIR Communities and other BMS RIs as resources from within the individual ELIXIR Nodes allows. This report provides an update to the ELIXIR Technical Services Roadmap (D4.1 and D4.2)[3] that were written at the end of PY1 and PY2 of the ELIXIR-EXCELERATE project. The ECP has also produced a public high-level description of the roadmap[4] based upon previous versions of this deliverable. The ELIXIR Technical Services Roadmap has been revised and updated to reflect the work and experiences that happened in PY3 and to help inform ELIXIR Researchers and Application Developers about the ECP services that will be available to them in the coming years. For ELIXIR Nodes and associated Infrastructure Service providers, this report identifies the technologies that should be deployed that will enable ELIXIR to provide a consistent set of Infrastructure Services to support life-science research in Europe.

It is important to note that the role of ELIXIR Compute experts is not to undertake middleware development within the ELIXIR-EXCELERATE project. Instead the focus is on leveraging the investment that has already been made nationally, by EC projects or

---

[3] https://drive.google.com/file/d/0B7btK9HAXhx1ZW9UTFZ1VDlfSTA/view
[4] http://drive.google.com/file/d/0B0KXZdVao0kqUE9BbXVrc3ZLY1E/view

commercially in services that can be integrated to serve the European research needs in biological information, and to influence on-going and future development priorities of European e-Infrastructures. The ECP has continued established strong collaborations with the European e-Infrastructure community through the EOSC Pilot, EOSC Hub, AARC/AARC2, and in the previous years EGI-Engage, EUDAT2020, AARC projects, and will establish Life Science AAI and Cloud leadership within the EOSC-Life project. EOSC - the European Open Science Cloud - represents a major new European initiative to integrate distributed resources to support Open Science. The EOSCpilot project is a community activity to help define what EOSC might be, and the EOSC-Hub project is tasked with implementing the initial phase of EOSC. The recently accepted EOSC-Life project brings together common requirements of the life science ESFRI's so that the European e-Infrastructure can support Life Science Research infrastructures more effectively. The role of the ECP is to help support and drive this process.

Even with the emergence of EOSC, the role of the ECP remains the same: to define a minimal 'neck' of an hourglass that ELIXIR Researchers and Application Developers can build upon and that ELIXIR Nodes and other infrastructure service providers can deploy services to provide.

The requirements defining the 'neck' of the hourglass have been broken down into individual Technical Use Cases (TUCs) that represent the 'generic' technical activities that need to be delivered by the ECP. The requirements from the various ELIXIR Communities and the supporting TUCs - initially coming from the Scientific Use Cases and Training activities in ELIXIR-EXCELERATE (WP6 to 9 and WP11) - are now documented elsewhere[5]. The focus of the ECP is therefore to identify the services that need to be deployed in the ELIXIR Nodes or drawn from European e-Infrastructures to meet these established needs.

The work undertaken during PY3 across the different tasks in WP4 of ELIXIR-EXCELERATE is summarised in Section 3 from a separate more detailed report[6]. This section provides an overview of the technical work and the collaborations that have been established with organisations outside the immediate scope of the project.

Section 4 provides an assessment as to the maturity of each TUC from being a Proof of Concept, an Emerging (or Prototype) Service, a Mature Service in production or a Legacy Service that is moving out of production use. Those TUCs where no progress has been made in PY3 are identified and the prioritisation of these TUCs in future years is flagged.

In Section 5, we assess ECP's readiness in supporting the different ELIXIR Communities by linking the maturity of the individual TUCs needed by the individual research communities. This assessment will help in driving the prioritisation of the TUCs and the

---

[5] https://docs.google.com/document/d/1mLeaFk5jlYlKQVRC6Vu2X7AKZBVGywFtf28t_NUO1iU/edit#

[6] https://docs.google.com/document/d/1uMCslVlaH1Lsne5F4nJfttUi1YSBmn7t5AQHaL2scj4/edit#

plans for during the remaining project timeline and transition into the ELIXIR 2019-23 programme.

Plans for PY4 are detailed in the 'Future Plans' section of the report.

To support the reader in reading this report a comprehensive glossary of technical terms is provided in Appendix A. More details of the individual TUCs and the ECP use cases coming from the ELIXIR Communities are provided in separate document.

2. ELIXIR-EXCELERATE Activities in Project Year 3

2.1 Background

During the third year of the ELIXIR-EXCELERATE project, WP4 has continued to build and develop the management, technical support, and technical structures needed for the ELIXIR Compute Platform (ECP) in response to the scientific and training use cases. Background for the work can be found in (D4.1 and D4.2)[7].

2.2 Leadership

The management and leadership of the ELIXIR Compute Platform (ECP) continues through an Executive Committee that meets twice per month to coordinate internal and external activity. The ECP task leads meet once a month and once a month there is a general ECP community meeting. Furthermore, the ECP participates in the monthly cross-platform meeting between the major ELIXIR platforms and ELIXIR communities involved in ELIXIR-Excelerate. The ELIXIR-Hub provides a dedicated Platform Coordinator.

Bi-directional technical interactions have continued with European e-Infrastructures (such as EGI, EUDAT and GÉANT). Most of the effort during PY3 has focused on projects related to the European Open Science Cloud (EOSC). The ECP plays a major role in EOSCpilot project started in January 2017 which through a series of three open calls now supports 4 life-science related projects[8] including PanCancer cloud analysis, analysis of life-science data sets, Cryo3D workflows and Bioimaging. The EOSC-Hub project started in January 2018 and the EOSC-Life project proposal which was coordinated by the ELIXIR-Hub has now been funded and will start in March 2019 with ECP partners having lead roles in WP5 AAI and WP7 cloud.

2.3 User Facing Support

User Facing Support with the users and relying parties of the ELIXIR AAI services has started. Direct interaction with the Scientific Use Cases and the Training activities coming from the ELIXIR Communities has triggered updates to the motivating use cases and the

---

[7] https://drive.google.com/file/d/0B7btK9HAXhx1ZW9UTFZ1VDlfSTA/view
[8] https://eoscpilot.eu/science-demonstrators

prioritised TUCs extracted from these use cases which are now reported separately[9]. In PY3 there has been a focus on plant and human data use cases with the ECP working closely with ELIXIR-EXCELERATE WP7 (Integrating Genomic and Phenotypic Data for Crop and Forest Plants) and WP9 (Human genetic Data) to support their use cases.

With the ECP now offering services to its users, the pilot support infrastructure needed to provide good user experiences (i.e. helpdesk, documentation, FAQs, etc.) has been established. A central contact point for the ELIXIR AAI has been identified and a team (1st line) to triage and assign support requests to the relevant service specific support teams (2nd line) is now being established.

2.4 Technical Infrastructure Integration

This work has been driven by the continuing work supporting the ELIXIR-EXCELERATE WP6 (Marine Metagenomics) use case that started in PY1, and expanded with support in PY2 for Human Data. The transfer of large volumes of electronic, confidential, human data coupled to cloud compute access to the data with the ELIXIR AAI is one of the key technology integrations achieved during PY3. Together WP4 and WP9 have worked to build Local EGA Core functionalities, that were demonstrated in May 2018. Local EGA has been an active collaboration with the Nordic Tryggve development team.[10] A technical demonstrator utilizing ELIXIR AAI, cloud and data transfers[11] has been prepared for the plants use case in PY3 as M4.3.

2.4.1 Authentication and Authorisation Infrastructure

The AAI service allows a user to create an ELIXIR identity based on a pre-existing identity (e.g. Google, ORCID or the researcher's home university as an attribute and an identity provider). Individual users can enroll into an ELIXIR Virtual Organisation, and potentially into groups within this Virtual Organisation (using Perun[12]). Group structures allow support of actual international research groups or projects, enabling resource allocation decisions to be recorded in the user and group metadata. This organisational information has then been exposed through the ELIXIR Proxy IdP to identified relying parties to use this information for authentication and authorisation decisions.

In addition to the ELIXIR-EXCELERATE funding, the AAI task has benefited from an ELIXIR IS (Link) in PY3, which has accelerated the development and integration of the AAI services needed to support human data. This includes step up authentication models through multi-factor authentication and bona fide researcher management service.

---

[9]

https://docs.google.com/document/d/1mLeaFk5jlYlKQVRC6Vu2X7AKZBVGywFtf28t_NUO1iU/edit#
[10] https://www.elixir-europe.org/events/elixir-webinar-transfer-large-volume-data
https://www.elixir-europe.org/events/webinar-access-to-sensitive-data-aai
https://www.youtube.com/watch?v=d8tIDZvDGKQ
[11] https://github.com/NBISweden/excelerate-demonstrator-4.3
[12] https://perun.cesnet.cz

The ELIXIR AAI has been driven primarily from the ELIXIR FI and ELIXIR CZ nodes with strong support and feedback from the ECP. Successful collaborations have been established with the AARC2 project in terms of providing requirements and benefiting from some of their service prototyping.

In the beginning of 2018 ELIXIR AAI enabled 493 institutions whose members can use ELIXIR AAI, integrated 67 production services in total and an additional 65 in testing, and had 1838 users and 370 groups.

2.4.2 Cloud & Compute

The support of the scientific (primarily marine metagenomics and human data) and training use cases by the Cloud and Compute Task has continued to be the main focus in PY3. ELIXIR Implementation Study support has been received for the Training use case, Workshop as a Service[13].

A recurring issue coming from both projects and their users of ELIXIR cloud resources is how these resources are provisioned as there is no ELIXIR wide mechanism for handling resource provisioning issues. Indeed, each cloud service provider operates independently meeting the needs of its local user community using the mechanisms defined by its funding agency. These mechanisms may include open access to funded researchers, formal peer review of proposed research use, pay per use, etc. External users, such as those from ELIXIR Nodes, may not be able to use these resources directly, but have to apply through a local collaborator. There is currently no clear mechanism as to how researchers who have no local cloud resources can gain access to cloud resources without paying for them directly, nor is there any mechanism likely to be put in place by which ELIXIR Nodes who might be able to provide such resources free at the point of delivery. The only foreseen mechanism is for usage to be reimbursed centrally for providing the service.

This is an issue that was a barrier to non-trivial resource use within EGI and similar issues are already occurring within the EOSCpilot - so ELIXIR is not alone in trying to establish a sustainable model and work will take place within the EOSC-Hub and EOSC-Life to explore and pilot different models.

The situation is complicated. The long-term aim of the ECP is to increase transparency of the details of infrastructure (resources) provisioning/allocation. With the support of e.g. EOSC-Life development ECP expects to be able to better report international consumption of compute and storage by the Life Science users for the purposes of data analysis.

The work has been primarily led by EMBL-EBI and ELIXIR CZ and resulted in a strong collaboration with other ELIXIR sites, in particular with ELIXIR FI, and with e-Infrastructure activities (such as EGI, Helix Nebula, and the European Open Science Cloud).

---

[13] http://bit.ly/ELIXIRWaaS

2.4.3 Storage and Data Transfer

Data transfers are needed across all the scientific use cases and various data transport mechanisms have been investigated to organise data transfers between ELIXIR data centres. FTS3, a file transfer coordination service, has been integrated with the ELIXIR AAI as a potential service for the ECP. The Reference Data Set Distribution Service being developed with the support of the EUDAT2020 project is expected to be available early in PY3 for use by the ECP and to enter production in 2018. This task has been supplemented during PY3 by an ELIXIR funded IS[14].

The work has been primarily led by ELIXIR SE, EMBL-EBI and ELIXIR NL.

2.4.4 Infrastructure Services Registry

An Infrastructure Service Registry is needed to provide a live picture of the available technical capabilities of the ECP. The information in the Infrastructure Services Registry is composed of both static information (e.g. contact URL, physical capacity) and dynamic capability information (e.g. free CPUs, free storage). Similar work is being undertaken by the EOSC-Hub project and we are waiting for this work to stabilise to see how it could be adopted by the ECP.

This work is being led by EMBL-EBI and ELIXIR CZ.

3. Assessment of Technical Use Cases

The following table provides an overview of the maturity status[15] of each TUC and how the maturity has changed in recent years.

Table 1. Assessment of the overall progress on individual TUCs over PY3

ID	Technical Use Case	Status - PY3	Status - PY2	Status - PY1
1	Federated ID	Mature	Mature	Emerging
2	Other ID	Mature	Mature	Emerging
3	ELIXIR Identity	Mature	Mature	Emerging
4	Cloud IaaS Services	Complete	Complete	Complete[16]
7	Network File Storage	PoC	Later	PY2
8	File Transfer	Emerging	Emerging	PoC

---

[14] https://docs.google.com/document/d/1WaSxtIIMPVKeCPxlhMjA9eQWUjmYByJJOgeKAI_RmmY/edit
[15] TUC Status: Proof of Concept (PoC), Emerging Services (Prototype), Mature Services (Production), Legacy Services

[16] A list of available ELIXIR Clouds IaaS services was established.

9	Infrastructure Service Directory	Emerging	Emerging	PoC
10	Credential Translation	PoC	PoC	PoC
11	Service Access Management	Mature	Mature	Mature
12	Virtual Machine Library	Mature	Emerging	PoC
13	Container Library	PoC	PoC	PoC
15	Data Set Replication	Prototype	PoC	PoC
17	Endorsed Personal Data or Compute Access Management	Emerging	PoC	PoC
19	PID and Metadata Registry	PoC	Later	PY2
20	Federated Cloud IaaS	Emerging	Emerging	PoC
21	Operational Integration	Emerging	Emerging	PoC
22	Resource Accounting	Emerging	Emerging	PoC

Implementation has not yet started on all of the TUCs identified through the analysis of the ELIXIR-EXCELERATE scientific use cases. As indicated below, some of these TUCs are expected to be addressed in future project years (PY) of the ELIXIR-EXCELERATE project. Other TUCs (e.g. 5 & 14) were identified during the use cases analysis, but at present there is little evidence that these are a priority at this stage. This prioritisation is always subject to change in response to user feedback.

Table 2. Identified TUCs for later implementation

ID	Technical Use Case	Status PY3	Status PY2	Timeline
5	HTC/HPC Cluster	Later	Later	PY3/4
6	PRACE Cluster	Later	Later	Awaiting user demand
14	Module Library	Later	Later	Awaiting user demand
16	Infrastructure Service Registry	Later	Later	PY4 - Awaiting EOSC
18	Cloud Storage	Later	Later	PY3/4
23	Federated HTC/HPC Cluster	Later	Later	PY3/4

4. Supporting ELIXIR Communities

Given the TUC prioritisation coming from the motivating use cases and the work that has been undertaken in PY1 and PY2 by the ECP to implement the TUCs, the priorities for PY3 can be assessed. The following table shows that the ECP is capable of supporting the motivating use cases with TUCs classed as 'Emerging' or 'Mature'.

Table 3. List of services available or planned to be offered to different use cases and their plans for their use.

ID	Technical Use Case	Status - PY3	WP6	WP7	WP8+9	WP11
1	Federated ID	Mature	Y	Y		Y
2	Other ID	Mature				Y
3	ELIXIR ID	Mature			Y	Y
4	Cloud IaaS Services	Complete	Y	Y	Y	Y
5	HTC/HPC Cluster	Later - PY3/4	Y	Y	Y	Y
7	Network File Storage	PoC	Y	Y	Y	Y
8	File Transfer	Emerging	Y	Y	Y	Y
9	Infrastructure Service Directory	Emerging		Y		
12	Virtual Machine Library	Mature	Y		Y	Y
16	Infrastructure Service Registry	Later - PY3		Y	Y	
17	Endorsed Personal Data or Compute Access Management	Emerging			Y	
19	PID and Metadata Registry	PoC		Y	Y	Y
22	Resource Accounting	Emerging	Y		Y	Y
23	Federated HTC/HPC Clusters	Later				Y

5. Plans for Next Year

In PY4 (the final year of the ELIXIR-EXCELERATE project), the key challenge for the ECP is to start the move into establishing a safe, secure, trusted and reliable operational environment for life-science researchers from across Europe. Of particular focus will be:

- **Support**: The ECP will continue to provide the best possible support to the use cases and other activities as resources allow (e.g. the ELIXIR Beacon project, hosting plant genomics data, etc.). The central helpdesk established in PY3 will provide a route for direct support for specific technical issues and will be evolved in PY4 to provide support beyond AAI. This will be alongside the continued ongoing investment in building collaborative relationships between organisations, teams, nodes and use cases.
- **Service Delivery**: Discussions will continue within ELIXIR as to how the community expects ECP's services to be operated, sustained and potentially

retired using the service maturity model. Some services will be delivered through partnerships (as ELIXIR does not have the expertise and budget to run the services directly) while others may need to be delivered through ELIXIR Nodes (e.g. cloud/compute resources) as this provides the best sustainability route.

● **Human Data**: Building the trust between data owners and service providers through authorising computational access to key human datasets. ELIXIR has technical components towards a solution: the EGA (with central and local components), secure (hybrid) clouds, secure data transfers, and federated authentication and authorization infrastructure services.

● **Analysis Platform**: An ELIXIR funded IS is exploring the use of GA4GH-endorsed specifications from the Cloud Work Stream could play in accessing cloud and compute resources. The current demonstrator, which links into work from other ELIXIR Platforms provides and environment for running workflows, defining tasks, executing tasks and accessing data that improve the portability of analysis pipelines and lowering the barriers for researchers undertaking large scale analysis.

● **Containers as a Service**: To support container based workload the ECP will work with service providers within the ELIXIR Nodes and elsewhere in PY4 to prototype a container deployment and execution environment. This is a recognition that containers are becoming a first class citizen in running workloads alongside virtual machines. This container environment will support the GA4GH analysis platform being assembled by the ECP, workloads from the Tools platform, platforms such as Galaxy and users wishing to run containers natively.

● **European Open Science Cloud**: As the new European e-Infrastructure and Research Infrastructure integration point, the ECP will be active within EOSC through the ELIXIR Competency Centre in EOSC-Hub and will be supporting various use cases through EOSC-Life project - notably in AAI and cloud.

● **Technology Collaboration**: The ECP will continue to monitor the landscape for technologies (e.g. CWL, FTS3) that need to be supported on ELIXIR Nodes to deliver on the identified TUCs through a network of projects (e.g. EOSC-Hub, EOSC-Pilot, EOSC-Life, AARC2). Part of this work will take place through proposed collaborations with other ELIXIR Platforms: Human Data Platform with the ELIXIR Beacon network, Interoperability Platform through containerised CWL workflows, and the Tools Platform through a container registry.

● **Infrastructure Service Development**: With limited budgets within ELIXIR-EXCELERATE for service development, the ECP, in collaboration with and support from ELIXIR-Hub will continue the operation and further specific development of ELIXIR AAI, in collaboration with AARC2 and e-infrastructures on the development of the LS AAI and with EOSC linked projects on the development of the Reference Data Set Distribution Service (RDSDS) and GA4GH analysis platform.

● **Commercial Clouds**: Users of the ECP are able to access commercial cloud resources (at no cost to the end-user) through the Helix Nebula Science Cloud by using the ELIXIR AAI. This work uses a resource allocation purchased by EMBL

as a partner in the project and similar models will be explored within the EOSC-Life project.

- **Sustainability**: Sustainability of the ECP needs to focus on two aspects: sustainability of the service operation and sustainability of the service consumption. The operation of services affiliated to individual ELIXIR nodes are generally expected to be sustained by the node, however the consumption of these services (e.g. cloud compute or storage) especially by users from outside these nodes will need to be sustained if this usage model is expected to continue. Integration of an existing accounting model (e.g. EGI) into ELIXIR Cloud resources will help collect usage information across distributed sites. Sustainability of services delivered by single ELIXIR Nodes on behalf of the whole ECP (e.g. AAI) will need to be sustained through the Commissioned Service model or by partnership with another organisation (e.g. EOSC).

6. Appendix A: Glossary of Key Terms

AAI = Authentication and Authorisation Infrastructure. Processes to verify person who they claim to be and permit to do what they want to do.

AARC = Authentication and Authorisation for Research and Collaboration project.
More information at https://aarc-project.eu/

Ansible = A tool for software remote management. It can be used for software installation, configuration and other necessary actions found on its playbooks.

Availability = Availability is the ratio of time a system or component is functional to the total time it is required or expected to function. This can be expressed as a direct proportion (for example, 9/10 or 0.9) or as a percentage (for example, 90%).

Container = A container middleware is a virtualization layer between the application and the operating system. The containers isolate the runtime environment and allow distribution in the containers. The containers are more lightweight with less overhead than the virtual machine images as they do not include operating system.

CWL = Common Workflow Language: open standards for the description of computational data analysis workflows in a portable and executable manner.

Data Provider = the individual researcher or investigator or body of researchers or investigators that makes data available or submits data for access and use in the context of an ELIXIR Service.

DevOps, Development and Operations = Agile working method to develop eServices. Close cooperation on development and production.

ECP = ELIXIR Compute Platform.

eduGAIN = GÉANT's service that enables trustworthy exchange of information related to identity, authentication and authorisation (AAI).

EGA = European Genome-phenome Archive. The EGA provides a service for the permanent archiving and distribution of personally identifiable genetic and phenotypic data resulting from biomedical research projects. Data Access Committees (DACs) control the access policies. More information at https://www.ebi.ac.uk/ega/home

EGI = European Grid Infrastructure. A federation of shared computing, storage and data services from national and intergovernmental service providers that delivers sustainable, integrated and secure distributed computing services to European researchers and their international partners. More information at https://www.egi.eu/

ELIXIR Service(s) = refers to ELIXIR Services as defined in the Node Collaboration Agreements, i.e. Node-funded Services or Commissioned Services.

EOSC = European Open Science Cloud. An initiative being driven by the European Commission through the work initially of a High-Level Expert group[17] which has helped to define the scope and direction of the work. The vision is being refined within the community through the EOSCpilot project[18] and will be implemented through the EOSC-Hub project due to start in January 2018. The EOSC-Life project supporting science demonstrators from across the BMS RIs is due to start in March 2019.

Federation = different computing services and/or infrastructures adhering to a certain standard of operation in a collective manner to facilitate its communication and interoperability.

Galaxy = open source, web-based platform for data intensive biomedical research. More information at https://usegalaxy.org/

GÉANT = Interconnects NRENs in Europe. Various services such as identity federation interconnection service eduGAIN. More information at http://www.geant.org/

GitHub = Git is a version control system and GitHub is a service for git based projects. It allows public and private repositories (license costs). GitLab can be used to run a private instance.

GoCDB = EGI's Grid Configuration Database. Contains general information about the sites participating to the production Grid.

---

[17] https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud
[18] https://eoscpilot.eu/

GridFTP = High-performance data transfer protocol. Integrated to Grid Security Infrastructure.

HPC/HTC = High Performance Computing, High Throughput Computing. In HTC the tasks are loosely-coupled when as HPC task require low latency and high performance requirements.

IaaS = Infrastructure as a Service, infrastructure level cloud service where the user administer their virtualised hardware such as virtual machines and their network and storage.

IdP = Identity Provider. In addition to the identifier of the user also other user information may be delivered for the Service Provider (SP).

Image = A Virtual Machine image contains operating system and possible other software readily installed. An image is a file with specific format such as raw or qcow2. A conversion might be possible.

IS = Implementation Study. ELIXIR Implementation Studies provide a mechanism by which funds contributed by the ELIXIR Nodes can be used to explore issues around implementing a service. IS are put forward by an ELIXIR Platform to the Hub for consideration for 12-18 months funding. They are a lightweight precursor to possibly build up to a full ELIXIR Commissioned Service application.

Metadata = Metadata contains descriptive, contextual and provenance assertions about the properties of a Digital Object. Makes data findable, usable and documented. Minimally the PID.

NREN = A National Research and Education Network. Provides various level network services.

OpenStack = A cloud middleware to manage the virtualised hardware.

ORCID = Persistent digital identifier for researchers. More information at http://orcid.org/

PaaS = Platform as a Service, readily installed software such as application server to run or develop the applications.

Perun = Identity and access management system developed and run by CESNET. More information at https://perun.cesnet.cz/

PID = A persistent identifier is a long-lasting ID represented by a string that uniquely points to a digital object and that is intended to be persistently resolvable. Used in search, linking and identifying.

Pipelines = A set of data processing elements that are connected in series, where the output of one element is the input of the next one.The elements of a pipeline are often executed in parallel, so several processes happen at the same time and the final result is obtained combining the results of the different processes or stages.

PRACE = Partnership for Advanced Computing in Europe. More information at http://www.prace-ri.eu/

PY = Project Year.

Reliability = refers to the ability of a computer-related hardware or software component to consistently perform according to its specifications. In theory, a reliable product is totally free of technical errors.

Relying parties = refers to a server providing access to a secure software application, or a web site or other entity on the Internet that uses an identity provider to authenticate a user that wants to log-in.

SaaS = Software as a Service, service such as Google Docs. No need to install or administrator any software by the end user.

TUC = Technical Use Case has been defined by the ELIXIR Compute Platform to capture a technical capability that may be repeated (in slightly modified forms) across a number of Scientific Use Cases.

Virtualisation = Layer on top of the physical hardware to allow multiple users to utilise the hardware in a secure manner.

Virtual Machine = Server on top of the virtualisation layer with (guest) operating system which the owner of the virtual machine administrates.