EXCELERATE Deliverable 4.2

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:	REPORT: Updated ELIXIR Technical Services Roadmap	
WP No.	4	
Lead Beneficiary:	35 - MU	
WP Title	Tools Interoperability and Service Registry	
Contractual delivery date:	30 July 2017	
Actual delivery date:	31 August 2017	
WP leader:	Tommi Nyrönen (FI) and Ludek Matyska (CZ)	CSC, MU

Authors and contributors:

Lars Ailo Bongo (UIT - NO), Mikael Borg (BILS - SE), Amelie Cornelis (EMBL-EBI), Montserrat Gonzalez (ELIXIR), Luis Gracia (EMBL-EBI), Rob Hooft (DTL - NL), Jarno Laitinen (CSC - FI), Ilkka Lappalainen (CSC - FI), Mikael Linden (CSC - FI), Ludek Matyska (MU - CZ), Steven Newhouse (EMBL-EBI), Tommi Nyrönen (CSC - FI), Michael Prochazka (MU - CZ), Mirek Ruda (CESNET - CZ), Harri Salminen (CSC - FI), Christine Staiger (SURFsara - NL),

Table of content

1. Executive Summary

This version of the ELIXIR Technical Services Roadmap (ELIXIR-EXCELERATE deliverable D4.2) has been written at the end of PY2 of the ELIXIR-EXCELERATE project (August 2017) and represents an update to the previous version that was established in August 2016 (Deliverable D4.1). It will be updated again in Summer 2018 following the 3rd Annual meeting of the ELIXIR Compute Platform (ECP). The ELIXIR Technical Services 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.

This version contains information relating to the technical work undertaken primarily within the ELIXIR Compute Platform during PY2, an assessment of the Platform's capability that was 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 user-driven use cases. Overall, the ELIXIR Compute Platform will need to support these TUCs directly through services hosted by the ELIXIR Nodes or indirectly in partnership with European e-Infrastructure providers. These integrated services - the ELIXIR Compute Platform - will be available to bioinformatics experts for use to analyse globally significant data resources.

Progress in PY2 has been extensive with a range of services now becoming available for end-users to consider adopting. The ELIXIR AAI[1] 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) and for that individual to be enrolled into the ELIXIR Virtual Organisation, and potentially into groups within this Virtual Organisation (using Perun[2]). 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. These include the ELIXIR Intranet, other ELIXIR services, the EGI AAI Gateway to access EGI services. The expansion of the relying service provider network using the ELIXIR identity continues with commercial cloud service providers within the Helix Nebula Science Cloud project, and integration with EUDAT's B2ACCESS service which would allow access to EUDAT services.

Moving files (i.e. data) between sites is a key capability for the ECP. GridFTP servers have been deployed and integrated with the ELIXIR AAI and a 'heartbeat' service makes regular file transfers between 9 GridFTP services to provide a record of reliability and network performance. The Reference Data Set Distribution Service that is being developed in collaboration with the EUDAT2020 project will be introduced into the ECP in PY3 with the support of an ELIXIR Implementation Study.

Work within the ECP to integrate the cloud resources affiliated to the ELIXIR Nodes continues. The EGI Federated Cloud model continues to be evaluated by the ECP,

---

[1]

especially within the context of the emerging European Open Science Cloud (EOSC) initiative, where it may be used as the federation model. The ECP is engaging with EOSC through an ELIXIR Competency Centre that will be funded as part of the EOSC-Hub project which builds upon the work that has been undertaken in the EGI-Engage project. Using the EGI services to monitor ELIXIR services and record ELIXIR usage has been delayed while the impact of adopting the EGI model is fully assessed.

Technical discussions with the ELIXIR-EXCELERATE Use Cases (the four scientific use cases and the training activities) continue as the effort undertaken with Marine Metagenomics in PY1 is consolidated, and the focus in PY2 on Human Data starts to yield results. The PY2 activity started with a joint workshop between WP4 and WP9 (September 2017, EMBL-EBI, Hinxton, UK) to identify how the key issues (i.e. step-up authentication, secure data transport and secure cloud compute) could be achieved during PY2. Results will be presented via a webinar in Autumn 2017.

As the midpoint of the ELIXIR-EXCELERATE project is reached, the basic infrastructure of the ECP has been defined and the work needed to see how these components can be integrated, optimised and sustained to support European life-science research has begun.

2. 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	

3. Delivery and schedule

The delivery is delayed: ☑Yes No

4. Adjustments made

N/A

5. Background information

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

Work package number	WP4	Lead beneficiary	35 - MU
Work package title	Technical Services		
Start Month	1	**End Month**	48
Work Package Lead	Tommi Nyrönen (FI) and Ludek Matyska (CZ)		

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.

Description of work and role of partners

WP4 - Technical Services [Months: 1-48]
MU, EMBL, NBIC, UPF, CSC, UiT, SIB, CNRS, IP, LIU, CESNET, DTU, ATHENA RIC
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 ELIXIREXCELERATEscientific 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.

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 projects like 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) EXCELERA TE 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

Partner number, short name and effort: 1- EMBL 34.00; 6 - NBIC 2.00, SARA 13.00, RUG 2.00; 14 - UPF 23.50; 20 - CSC 48.00; 23 - UiT 2.00; 25 - SIB 2.00; 26 - CNRS 12.00; 29 - IP 6.00; 31 - LIU 12.00; 35 - MU 40.00; 36 - CESNET 24.00; 38 - DTU 6.00; 41 - ATHENA RIC 8.00

6. REPORT: Updated ELIXIR Technical Services Roadmap

6.1 Summary

Key ELIXIR Compute Platform achievements during Project Year 2 of ELIXIR-EXCELERATE

Collaboration within the ELIXIR Compute Platform
- Establishing the ELIXIR Identity through the ELIXIR AAI Proxy
- Expanding to 9 GridFTP services that are part of the data transfer 'heartbeat'
- Creating a public description of the platform roadmap
- Planning the expansion of the helpdesk to provide support for selected ECP services

Collaboration within the ELIXIR-EXCELERATE project
- WP6 to deploy the marine metagenomics data analysis software environment on ELIXIR cloud resources
- WP9 to demonstrate a PoC in PY2 of the authentication, authorisation, movement and analysis of human data on ELIXIR cloud resources
- WP11 to support use of clouds for bioinformatics training

Collaboration with European e-Infrastructures projects
- Providing input to AARC on the needs of ELIXIR for a European AAI for Researchers
- Deployment of the marine metagenomics use case (WP6) on EGI Federated Cloud through the ELIXIR Competence Center in the EGI-Engage project
- Providing feedback on EGI's Service Level Agreements and Operational Policies to make them suitable for ELIXIR
- Integrating the ELIXIR AAI into Helix Nebula Science Cloud project to the support the ELIXIR use cases
- Development of a Reference Data Set Distribution Service for deployment within ELIXIR with the support of the EUDAT2020 project

6.2.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 2014-2018 and beyond. Within the ELIXIR-EXCELERATE project the requirements coming from the ELIXIR research community are encapsulated in the Scientific Use Cases and Training activities (WP6 to 9 and WP11). These user-facing activities were broken down into individual Technical Use Cases (TUCs) that represented 'generic' technical activities that could be used to demonstrate elements of the ECP as detailed in Appendix B. The focus of the ECP is to identify the services that need to be deployed in the ELIXIR Nodes or drawn from European e-Infrastructures to meet these established needs.

This report provides an update to the ELIXIR Technical Services Roadmap (D4.1)[3] that was generated in Summer 2016 by the ECP at the end of PY1 of the ELIXIR-EXCELERATE project. The ELIXIR Technical Services Roadmap has been revised and updated to reflect the work and experiences that happened in PY2 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 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 EGI-Engage, EUDAT2020, AARC and EOSCpilot projects, and is building new collaborations within the EOSC-Hub 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.

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 and support.

The technical work undertaken during PY2 across all the TUCs is summarised in the next section and detailed in Appendix C. This section provides an overview of the work undertaken within WP4 of the ELIXIR-EXCELERATE project and the collaborations that have been established with organisations outside the immediate scope of the

---

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

project. In addition an assessment is made of the progress across each TUC. This assessment classifies the maturity of each TUC as 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 PY2 are identified and the prioritisation of these TUCs in future years is flagged.

The 'Motivating Use Cases' section describes the Scientific and Training Use Cases that are being used to drive the development of the ECP. This section has been updated to reflect the improved understanding the use cases have of how they will undertake their work, and the improved understanding the ECP has gained during PY2 of how the use cases will exploit the ECP services. For each use case the supporting TUCs have been identified and prioritised.

In the 'Supporting the Use Cases' section, an assessment is made of the TUCs needed to support each use case, and the maturity of the TUCs needed to support the researchers. This assessment will help in driving the prioritisation of the TUCs and the plans for PY3 which 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 are provided in Appendix B.

6.3 ELIXIR Compute Platform Activities in Project Year 2

Background

During the second 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 in response to the scientific and training use cases. More details of the work in PY2 are provided in Appendix C and details from PY1 can be found in D4.1.

Leadership

The management and leadership of the ELIXIR Compute Platform (ECP) continues through an Executive Committee that meets at least once every month to coordinate internal and external activity. This year has seen welcomed additional support from the ELIXIR-Hub which has provided funds to appoint a Platform Coordinator. However, due to staffing issues this post has only been filled at intervals during the year until June 2017, when a new Coordinator was appointed. While good bi-directional technical interactions have continued with European e-Infrastructures (such as EGI, EUDAT and GÉANT) significant future discussions have focused this year around the European Open Science Cloud (EOSC). This has taken place through the EOSCpilot project that started in January 2017 and contribution to the EOSC-Hub project preparation which was submitted in March 2017 and will now start as a funded activity in January 2018. ELIXIR has proposed many science demonstrators to the open call organised by the EOSCpilot and has been successful with the PanCancer analysis and the EGA data analysis demonstrators[4].

User Facing Support

User Facing Support with the users of the ECP has also started. Interaction with the Scientific Use Cases and the Training activities has triggered updates to the motivating use cases and the prioritised TUCs extracted from these use cases which are captured in this document - the ELIXIR Technical Services Roadmap. In PY2 there has been a focus on human data with the ECP working closely with ELIXIR-EXCELERATE WP9 (Marine Metagenomics) to support their use case. This will lead to a demonstrator towards the end of 2017. With the ECP now offering services to its users, the support infrastructure needed to provide good user experiences (i.e. helpdesk, documentation, FAQs, etc.) has been established. A central contact point 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.

---

[4] https://eoscpilot.eu/science-demonstrators

Technical Infrastructure Integration

The technical activities within the ECP - the implementation of the TUCs (which are listed in Tables 1 & 2 and detailed in Appendix B) - has progressed with work in the interlinked Authentication & Authorisation Infrastructure (AAI), Storage and Data Transfer, Cloud & Compute, and Infrastructure Services Registry activities continuing. This work has been driven by the continuing work supporting the ELIXIR-EXCELERATE WP9 (Marine Metagenomics) use case that started in PY1, and expanded with support in PY2 for Human Data.

Authentication and Authorisation Infrastructure

In addition to the ELIXIR-EXCELERATE funding, the AAI tasks has benefited from an ELIXIR Implementation Study (Link) in PY2, which has accelerated the development and integration of the AAI services needed to support human data. This includes step up authentication models. The need for a further Implementation Study has been identified for PY3 to assess which of the ELIXIR AAI services can be migrated to European e-Infrastructures such as EOSC or GEANT.

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 AARC & AARC2 projects in terms of providing requirements and benefiting from some of their service prototyping.

The ELIXIR AAI is currently a production service used by increasing number of service providers within the ELIXIR community, with three hundred identity providers, more than 800 users and tens of services connected.

Cloud & Compute

The support of the scientific (primarily marine metagenomics and human data) and training use cases by the Cloud and Compute Task has been the main focus in PY2. 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 could be compensated centrally for providing the service. The situation is made even more complicated if the long-term aim of the ECP is to deliberately hide the details of

infrastructure (resources) provisioning/allocation from the users. 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.

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).

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. A heartbeat service operates between 9 ELIXIR Nodes using GridFTP endpoints to provide a regular record as to the network performance between these nodes. FTS3, a file transfer coordination service, is being 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. A funded IS[5] will start on 1.10.2017 to support a number of ELIXIR nodes in the testing and integration of this service into production.

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

Infrastructure Services Registry

ELIXIR Infrastructure Service Registry will be deployed to provide a live picture of the technical capabilities of the ECP, and will complement other registries such as the ELIXIR-EXCELERATE WP1 Tools Registry (http://bio.tools). 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). Some prototyping has taken place during PY2 but the work will be focus in PY3 and PY4.

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

Table 1. Assessment of the overall progress on individual TUCs over PY2

ID	Technical Use Case	Status - PY2[6]	Status - PY1

---

[5]
https://docs.google.com/document/d/1O3fFCK5Gyar1ooKOcKGwicT2VY3Z80wFgIHC7xSFEcA/edit
[6]TUC Status: Proof of Concept (PoC), Emerging Services (Prototype), Mature Services (Production), Legacy Services

1	Federated ID	Mature	Emerging
2	Other ID	Mature	Emerging
3	ELIXIR Identity	Mature	Emerging
4	Cloud IaaS Services	Legacy	Complete[7]
8	File Transfer	Emerging	PoC
9	Infrastructure Service Directory	Emerging	PoC
10	Credential Translation	PoC	PoC
11	Service Access Management	Mature	Mature
12	Virtual Machine Library	Emerging	PoC
13	Container Library	PoC	PoC
15	Data Set Replication	PoC	PoC
17	Endorsed Personal Data or Compute Access Management	PoC	PoC
20	Federated Cloud IaaS	Emerging	PoC
21	Operational Integration	Emerging	PoC
22	Resource Accounting	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. 7 & 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 PY2	Timeline
5	HTC/HPC Cluster	Later	PY3/4
6	PRACE Cluster	Later	Awaiting user demand
7	Network File Storage	Later	PY2
14	Module Library	Later	Awaiting user demand

---

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

16	Infrastructure Service Registry	Later	PY2
18	Cloud Storage	Later	PY2
19	PID and Metadata Registry	Later	PY2
23	Federated HTC/HPC Cluster	Later	PY3/4

6.4 Motivating Use Cases for the ELIXIR Compute Platform

This section provides an overview of each of the life science research-driven use cases that are the current focus of the ELIXIR Compute Platform, and how scientific use cases have been broken down into specific individual Technical Use Cases. The immediate scope of the ECP within the ELIXIR-EXCELERATE project is around WP6-9 (Scientific Use Cases) and WP11 (Training).

This section has been reviewed and updated from D4.1 but included here in full as the use cases that the ECP needs to support are fundamental to defining the Technical Use Cases.

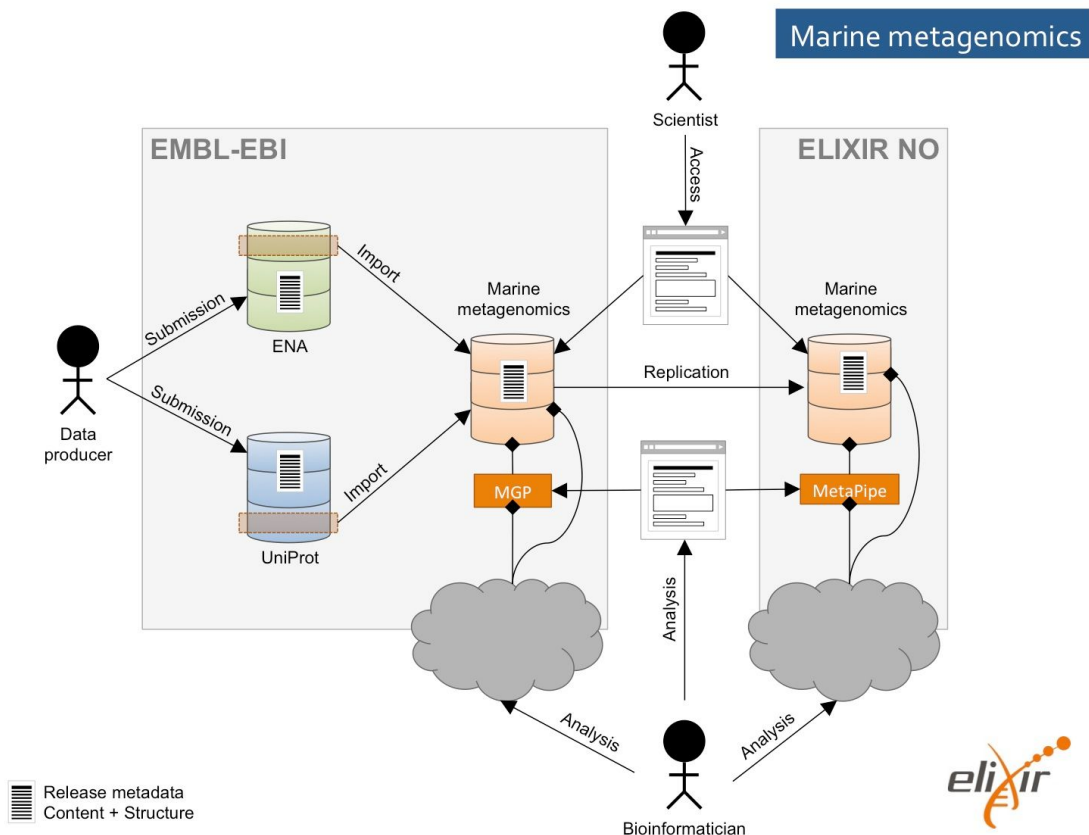## ELIXIR-EXCELERATE WP6 Use Case A: Marine Metagenomics

Figure 1. Example of Marine Metagenomics use case. The data producer uploads information to different databases, which can feed other databases or even be replicated in the different ELIXIR Nodes. The scientists can request access to the datasets in the databases. The

bioinformaticians can access datasets and process or analyze them using cloud or computing resources.

The marine metagenomic infrastructure provides a web based portal that will act as a driver for research and industrial innovation. The portal provides access to a marine reference database populated with data produced by running pipelines[8] on internal High Throughput Computing (HTC) clusters using data generated from within the collaboration, but also collected from other data sources. These pipelines are being ported from the current cluster to the future cloud infrastructure.

The technical solution and architecture delivered to support the marine metagenomics software workflows and the underlying user portal services and data processing was presented in the ELIXIR Compute webinar held in November 2016. The demonstrator is currently sustained with the support from ELIXIR (mainly CZ; EMBL-EBI, FI nodes) and with a European e-Infrastructure collaboration (EGI.eu).

Table 3. List of TUCs supporting the Marine Metagenomics use case

ID	Technical Use Case	Comment
1	Federated ID	This is needed to authenticate researchers to the portal. An ELIXIR ID (TUC 3) built on top of Federated (TUC 1) and Other IDs (TUC 2) would also meet this need. Identities may need to be collected into groups (TUC 11) to manage access to Infrastructure Services which would need to be integrated with the ELIXIR AAI. Any environment provisioned from VMs or containers might also need to be integrated with the ELIXIR AAI if it is to be generally accessible. These credentials may need to be translated to other systems (TUC 10).
5	HTC/HPC Cluster	An HPC cluster currently provides the analysis capacity to generate the data for the marine reference database. The Norwegian e-Infrastructure clusters (NOTUR) is currently being used to support this work. A Module Library (TUC 14) could be used to provide a common cluster environment.
4	Cloud IaaS Services	A Cloud IaaS is needed to provide the on-demand services for the user-driven analysis and the elasticity needed for processing the variously sized datasets for the reference database. For the web portal to be able to easily deploy the analysis pipelines and access different Cloud IaaS's consistently these services should be Federated (TUC 20) and integrated into an operational infrastructure (TUC 21) where infrastructure services can be declared within an Infrastructure Services Directory (TUC 9) and then discovered from an infrastructure service registry (TUC 16) and the usage can be accounted for (TUC 22). The pipelines are currently being ported to the EMBL-EBI Embassy cloud

---

[8] Definition available in the Glossary of terms.

		and CSC's cPouta and EGI Federated Cloud on ELIXIR-CZ's cloud resources. EGI-Engage ELIXIR Competence Center has provided support for using cPouta. CSC's Ansible playbooks (recipes) to setup a Spark data analysis virtual cluster has been made available. https://github.com/elixir-marine/mmg-cluster-setup
7	Network File Storage	Needed to provide temporary storage of the data as it is transferred from the primary sources (e.g. ENA and UniProt) to temporary storage space in the Marine Metagenomics infrastructure. This infrastructure may be a Federated HTC/HPC Cluster (TUC 23) or Cloud (TUC 20) including Cloud Storage (TUC 18).
8	File Transfer	Move the data from primary to temporary storage. This will include a small amount of data (<1TB) from the portal. Over 75 TB compressed data is estimated to be processed over the 4 years including data in various archives such as ENA (TUC 15).
22	Resource Accounting	Important so that once distributed computing and cloud resources are used a complete view of the consumed resources can be recorded and analysed.

ELIXIR-EXCELERATE_WP7 Use Case B: Genomic and Phenotypic Data for Crop and Forest Plants
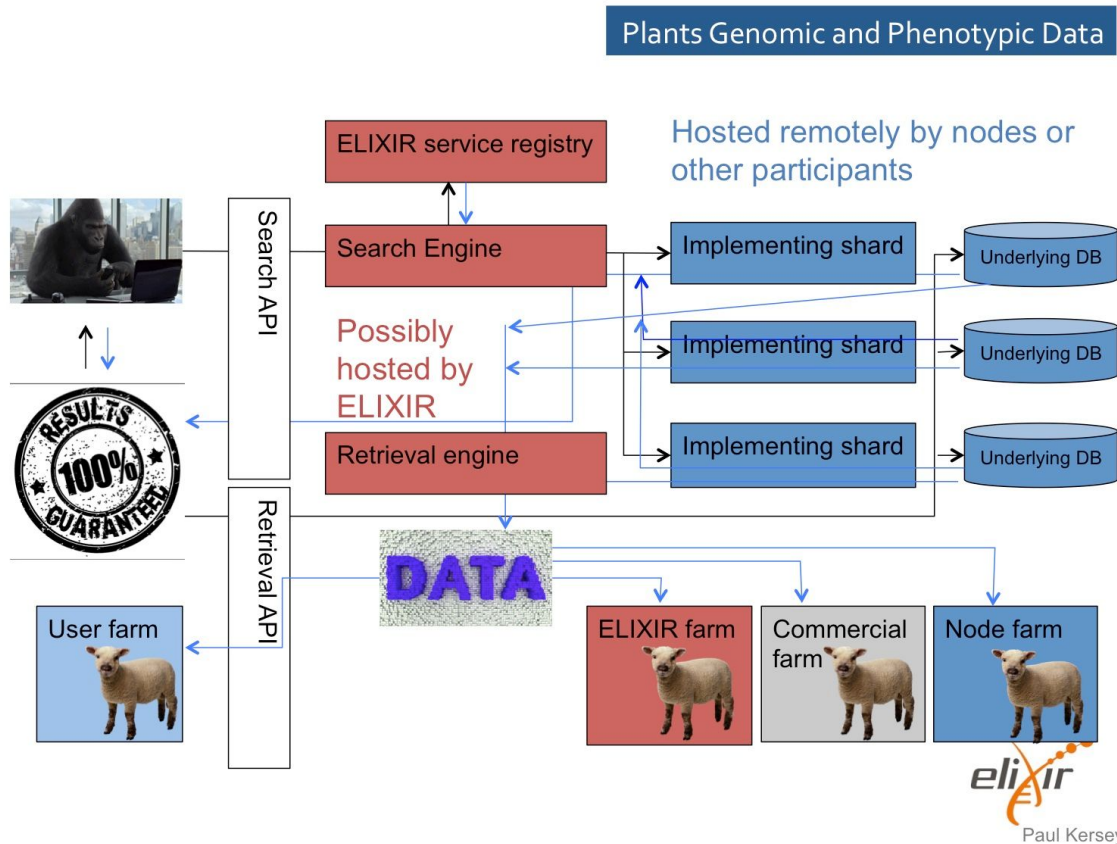

Figure 2. Example of Genomic and Phenotypic Data for Crop and Forest Plants use case. The data providers construct the Underlying DB and the implementing shard provides an index in a standard format for all the different databases. Since indexes follow the same structure they can be integrated and can be used by search engines to look for the data that the user needs. Any new database is registered in the ELIXIR service registry so its information is available.Once the search is completed, the information retrieved can be transferred to an ELIXIR, commercial, Node, or in the user computing facility and the results of the analysis are provided to the end user.

This scientific use case integrates Genomic and Phenotypic Data for Crop and Forest Plants from a variety of open access and open data sources. All of these data sources will conform to minimum standards and have agreed data discovery and retrieval mechanisms. The only 'central' component is a search engine that receives search requests from the users and passes integrated search results retrieved from the distributed data sources back to the user. Based on these results a user will want to select a cloud infrastructure that they have access to and transfer the selected data to that cloud resource to undertake their own analysis.

Therefore the use case has two aspects:

- Centralised Data Discovery: The Infrastructure Services Registry could be used to provide a platform for registering the data source URLs (which provide the search API) and the monitoring service (part of TUC 21) could identify which of the available data sources are conformant with the API. A searchable distributed PID and metadata registry (TUC 19) could also provide this data search capability.
- Analysis of Identified Data: Once the data that needs to be analysed has been identified, an existing cluster needs to be identified (TUC 5) or a new cluster provisioned on a cloud specifically to support this analysis (TUC 4). Storage needs to be provisioned and associated with the analysis resource (TUC 7) before the selected data is copied to this file space (TUC 8). The provisioned analysis environment (the cloud/cluster and the selected data) is then assigned to the user for their analysis to take place.

Table 4. List of TUCs supporting the Genomic and Phenotypic Data for Crop and Forest Plants

ID	Technical Use Case	Comment
4	Cloud IaaS Services	A cloud infrastructure is needed to support and scale the central open access search service and to provide on-demand consistent access to federated services (TUC 20) and a Virtual Machine Image library to hold pre-defined images for users to use (TUC 12). Both of these could be hosted by ELIXIR.
9	Infrastructure Service Directory	The ability to search for compliant and available Data Sources and to use the returned URLs for distributing search queries coming into the central search portal.
19	PID and Metadata Registry	Could provide the local data search infrastructure, but there may be more appropriate domain specific solutions.
16	Infrastructure Service Registry	Registry that holds the location of the conformant data discovery instances and the Cloud IaaS instances (and associated Network File Storage endpoints) that are available for use. These infrastructure services are integrated (TUC 21) and accounted for (TUC 22).
5	HTC/HPC Cluster	A HTC cluster is used to analyse data and these need to be federated (TUC 23) so that there is a consistent user experience whichever cluster is selected for use.
7	Network File Storage	Needed to provide temporary storage of the data as it is transferred from the distributed sources through the Retrieval API to the Cloud IaaS provider for the users analysis.
8	File Transfer	Move the data from primary to temporary storage at the selected Cloud IaaS provider. A PID and metadata registry (TUC 19) is used to keep track of the location and content of data sets.

1	Federated ID	Authentication is *not* needed (nor is it required by WP4) to search or retrieve the data - all the data is open access. However, it is needed to authenticate researchers to use the selected Cloud IaaS (TUC 4) or HTC/HPC Cluster (TUC 5) and to write the data into either Network File Storage (TUC 7) or Cloud Storage (TUC 18) from the selected open data sources. An ELIXIR ID (TUC 3) built on top of Federated and Other IDs (TUC 2) would also meet this need and credentials may need to be translated (TUC 10) to access the services.

ELIXIR-EXCELERATE WP8 Use Case C: Rare Disease & WP9 Use Case D: Human Genetic Data

These two Scientific Use Cases around Human Data were seen to have essentially identical technical requirements.

The WP8 Rare Disease Use Case is described below:
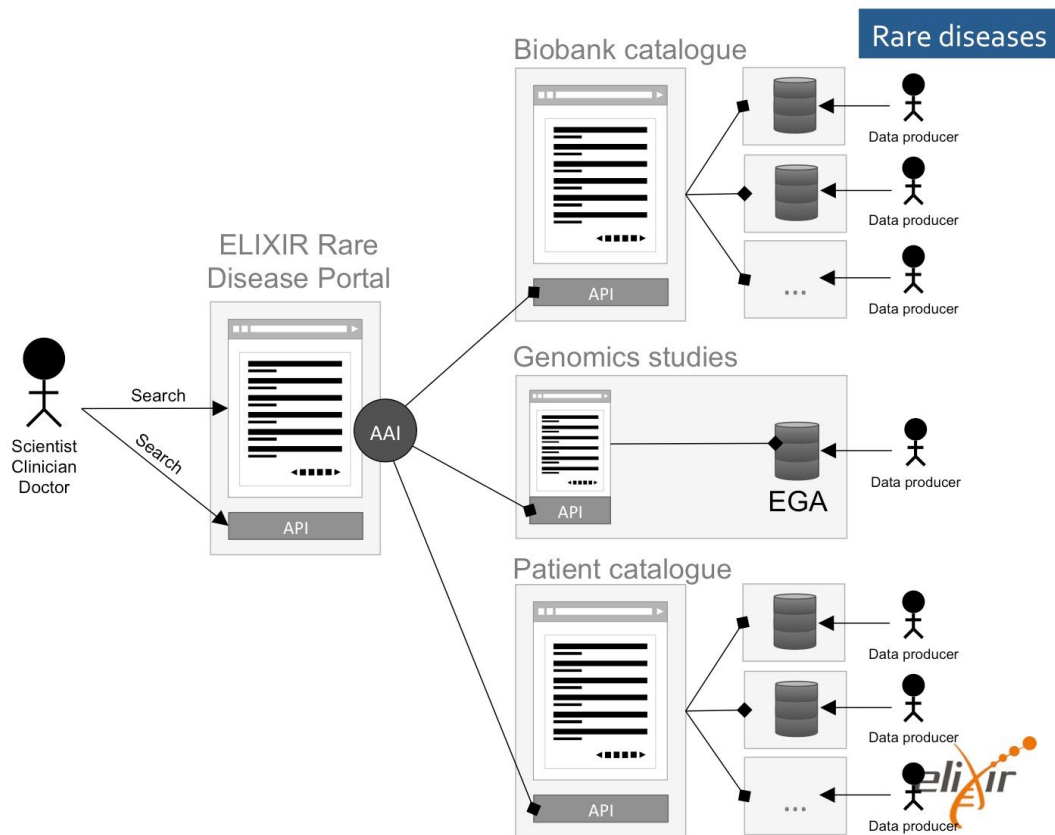
Figure 3. Example of Rare disease use case. The example considers three different information types: sample related information (indexed in Biobanks), genomic information (indexed as genomic studies using the EGA system), and the patient information (indexed in the patient catalogue). The three different sets of indexes are integrated in a upper level by the ELIXIR Rare disease portal. The end user can access to the ELIXIR portal and submit information request after being authenticated and authorised by the ELIXIR AAI.

This use case is about supporting research on rare (1 in 2000 people) chronic or genetic diseases that use EGA as data sources - access to which is controlled. The metadata on a patient (i.e. illness, treatments, outcomes), patient samples stored in a biobank, and any sequenced material stored in EGA is searchable through a central portal which can only be accessed by authorised users. The portal queries the individual national search engines on behalf of the users. Selected datasets can then be downloaded into an EGA-compatible cloud or cluster local to the researcher.

WP8 is expecting support from WP4 as to how to portably specify the pipeline setup for operation across the ECP. Security is an important concern beyond just accessing the compute and cloud resources as access to the data being processed by the pipelines needs to be strictly controlled.

The WP9 Rare Disease Use Case is described below:

The WP9 Human Genetic Data Use Case will use the ELIXIR framework for secure submission, archiving, dissemination and analysis of human access-controlled data. The work will extend and generalise the system of access authorisation management and high-volume secure data transfer developed in the EGA project to address the secure data access needs across ELIXIR resources. A centrally provided service will allow authorized third-party services to programmatically check that a user is authorised to access data stored in ELIXIR coordinated distributed repositories. This will also provide support for the dataset owners (e.g. usable technology, implementing policy, granting permissions).
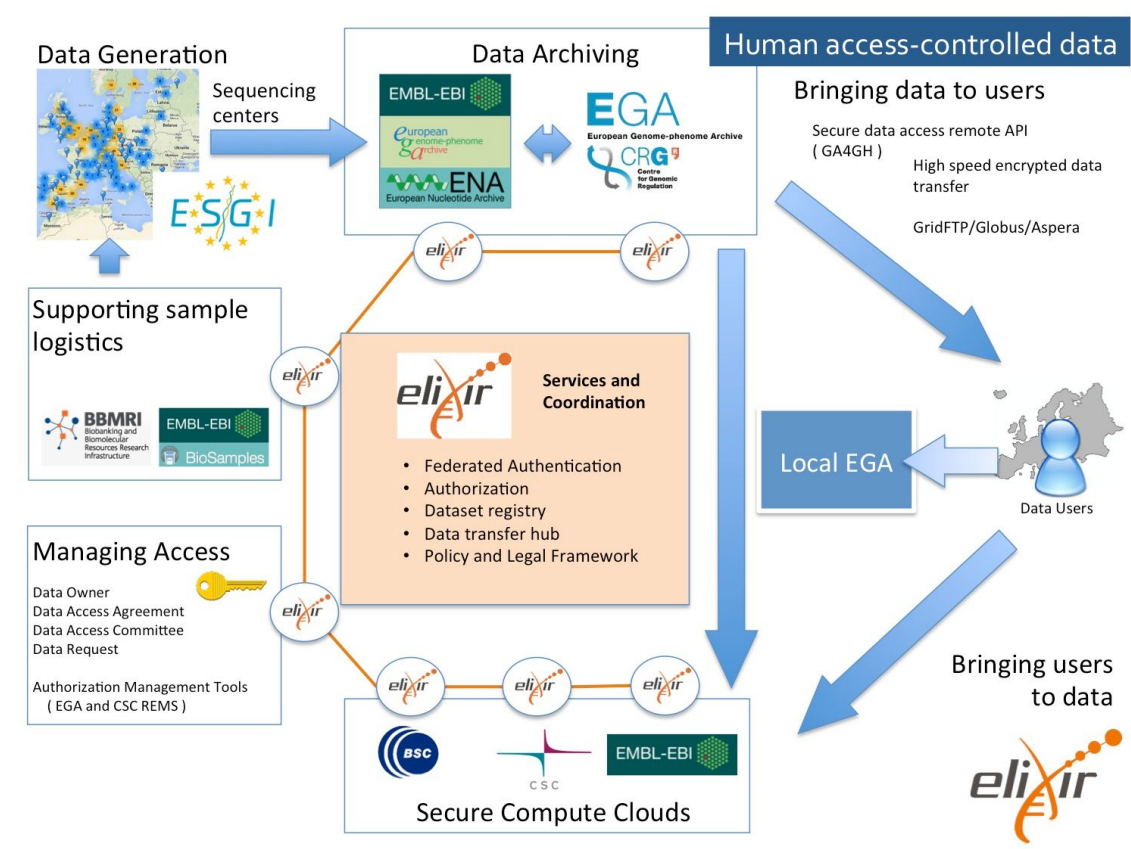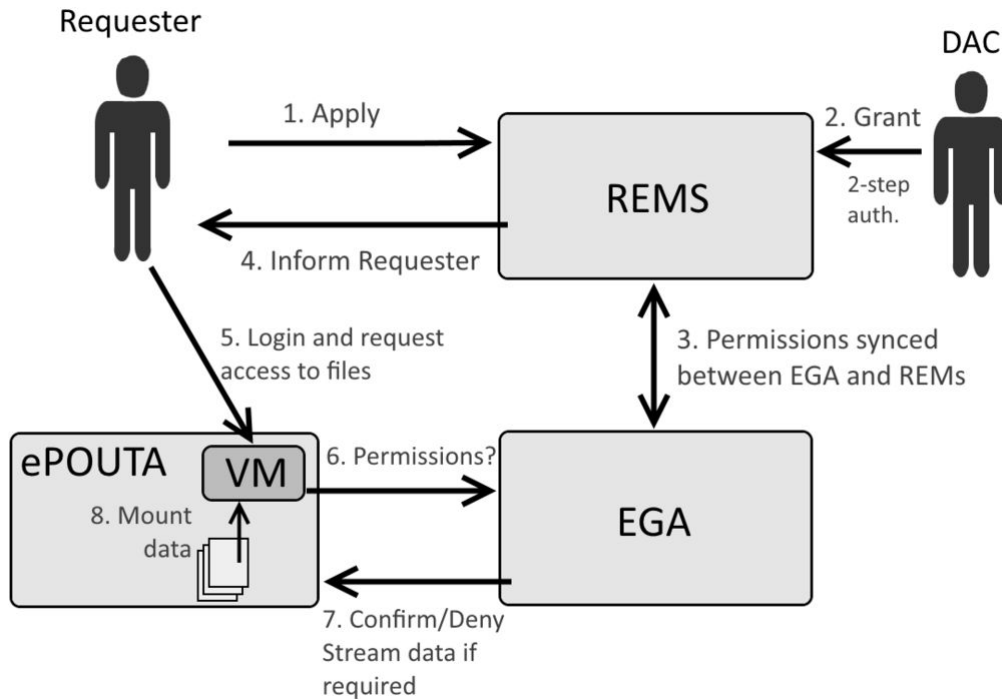
Figure 4. Example of Human data use case

Figure 5. ELIXIR-EXCELERATE WP9 integration with WP4 being developed in PY2 for the Human Data Use Case

Table 5. List of TUCs supporting the Rare disease and Human data use cases

ID	Technical Use Case	Comment
3	ELIXIR ID	This is needed to authenticate researchers to users accessing the central services or when accessing data or compute/cloud services. An ELIXIR ID built on top of Federated (TUC 1) and Other IDs (TUC 2) would also meet this need, which may include the ability to translate credentials (TUC 10).
4	Cloud IaaS Services	A cloud infrastructure (whose security processes have been accepted by EGA) is needed to undertake the data analysis. Cloud services need to be federated (TUC 20) to provide uniform operation (TUC 21) and secure access to storage (TUC 18) and the use of these services needs to be accounted for (TUC 22). Not all federated cloud services will necessarily be considered to be secure enough for use to analyse access-controlled data, i.e. only a subset may be available.
5	HTC/HPC Cluster	A secure cluster environment is needed to support the secure analysis of any downloaded data. Cluster services need to be federated (TUC 20) to provide uniform operation and secure access (TUC 21) and the use of these services needs to be accounted for (TUC 22).
7	Network File Storage	Needed to provide temporary secure storage of the data as it is transferred from EGA for analysis on a cluster or cloud.
8	File Transfer	Move the data from primary to temporary storage at the selected Cloud IaaS or HTC/HPC Cluster service provider. Data sets may be replicated (TUC 15) and pre-staged for popular data sets on popular service providers.
16	Infrastructure Service Registry	Holds the location of the Cluster, Cloud IaaS instances (and associated Network File Storage endpoints) that are available for use. The Infrastructure Services Directory (TUC 9) provides essential technical and contact details (e.g. security incident contact).
12	Virtual Machine Library	It is foreseen that endorsed Virtual Machine Images will be provided for use either in an EGA endorsed (for security purposes) ELIXIR Cloud service or in a local institutional cloud.
17	Endorsed Attribute Management	Individuals need to be authorised to access the various services by having their research activities approved by the appropriate Data Access Committee.
19	PID & Metadata	Used to provide a registry of patients (authorised users only), biobank samples (some of which will have been sequenced in EGA) and the location of registered data sets and any replicas that might be pre-staged onto the secured cloud services.

ELIXIR-EXCELERATE WP11 Use Case E: Training

The training Use Case has prioritised the identity and cloud (IaaS) TUCs (1-4). More detailed descriptions of the training requirements have been captured[9] and have been summarised in the following table.

Table 6. List of TUCs supporting the Training use case

ID	Technical Use Case	Comment
1	Federated ID	Various methods for trainee and trainer authentication are needed. For training of researchers Federated ID + Other ID (Google) is expected. When ELIXIR ID is established, it will be used in addition to Federated ID.
2	Other ID	Useful for newcomers to training session, without any pre-registration.
3	ELIXIR ID	This is needed to authenticate trainers to accessing the central services, training portals, internal servers, etc.
4	Cloud IaaS Services	A cloud infrastructure is needed to run e-learning, training courses etc. The number of VMs needed for each course depends on the course topic and the number of students.
5	HTC/HPC Cluster	Only if training HTC/HPC infrastructure is required to perform the training exercises.
7	Network File Storage	Similar to TUC 18, shared data for training.
8	File Transfer	Move the data from home institute to training environment. To download VM image from library to training environment.
12	Virtual Machine Library	It is foreseen that training environment will use VM images, where software which is trained is pre-installed. Galaxy and Chipster images are the first examples.
13	Container Library	Similar/alternative for TUC 12, software pre-installed in Docker images is expected pattern.
18	Cloud storage	Useful for sharing data between VMs (referenced data or pre-installed tools in Galaxy/Chipster)
22	Resource	Expected to use for user charging and for resource planning for training.

---

[9]

https://docs.google.com/spreadsheets/d/1fCcbbOpNHg9TGmSHxBDowDsnHN7YHhmq79lj5fPL_Gk/edit#

	Accounting	
23	Federated HTC/HPC Cluster	Training is expected to run on one site, with other sites to be ready as replacement/overflow in case of problems.

Other ELIXIR Use Cases

The ELIXIR Beacon project has been part of the ELIXIR Implementation Study 2016 with the aim to initiate a GA4GH Beacon API in a number of partnering ELIXIR Nodes by the end of 2016. The Beacon service initially provided a simple query interface for genome information. The ECP improved that interface by integrating the required authentication and authorization tools to the Beacon prototype implementation. The work started in June 2016 and continued through ELIXIR-EXCELERATE PY2.

Supporting the Use Cases

Given the TUC prioritisation coming from the motivating use cases and the work that has been undertaken in PY1 by the ELIXIR Compute Platform to implement the TUCs, the priorities for PY2 can be assessed.

The following table shows that a basic ELIXIR Compute Platform capable of supporting the motivating use cases is beginning to emerge. Services classed as 'Proof of Concept' or 'Emerging' are available across many of the key TUCs.

Table 7. List of services available or planned to be offered to different use cases and their plans for their use.

ID	Technical Use Case	Status - PY2	Status - PY1	WP6	WP7	WP8+9	WP11
1	Federated ID	Mature	Emerging	Y	Y		Y
2	Other ID	Mature	Emerging				Y
3	ELIXIR ID	Mature	Emerging			Y	Y
4	Cloud IaaS Services	Complete	Complete	Y	Y	Y	Y
5	HTC/HPC Cluster	Later - PY3/4	Later - PY3/4	Y	Y	Y	Y
7	Network File Storage	Later - PY3	Later - PY2	Y	Y	Y	Y
8	File Transfer	Emerging	PoC	Y	Y	Y	Y
9	Infrastructure Service Directory	Emerging	PoC		Y		
12	Virtual Machine Library	Emerging	PoC	Y		Y	Y
16	Infrastructure Service Registry	Later - PY3	Later - PY2		Y	Y	
17	Endorsed Personal Data or Compute Access Management	PoC	PoC			Y	
19	PID and Metadata Registry	PoC	Later		Y	Y	Y
22	Resource Accounting	Emerging	PoC	Y		Y	Y
23	Federated HTC/HPC Clusters	Later	Later				Y

Plans for Next Year

In PY3 (the start of the second half 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 will provide a route for direct support for specific technical issues and will be evolved in PY3 to meet these needs. 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. In PY3 the ECP and Training platform will jointly create a process as to how virtualized servers can be delivered for trainers within ELIXIR.

- **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 and federated authentication and authorization infrastructure services. Development of these solutions will continue with an ELIXIR funded IS on the ECP's AAI services and a collaborative IS with the ELIXIR Human Data platform on the GA4GH Beacon network.

- **Analysis Platform**: Exploring through an ELIXIR funded IS the role that a GA4GH-endorsed platform could play in accessing cloud and compute resources. GA4GH's cloud working group are defining a suite of APIs around running workflows, defining tasks, executing tasks and accessing data that improve the portability of analysis pipelines and lowering the barriers to entry.

- **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 and will be providing requirements for other services, notably AAI, which EOSC could over time provide to the ELIXIR community.

- **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, AARC2, EUDAT2020, EGI-Engage). 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 has been collaborating with AARC on the development of the ELIXIR AAI and with EUDAT2020 on the development of the Reference Data Set Distribution Service (RDSDS). These collaborations will continue into PY3. With the rollout of RDSDS within the ECP supported by an ELIXIR funded implementation study for 6 months which will also see the integration of FTS3 with the ELIXIR AAI. Following a Proof of Concept developed in PY2, the ELIXIR AAI will be integrated with B2ACCESS (the EUDAT AAI) allowing ELIXIR users to authenticate to EUDAT services.
- **Commercial Clouds**: ECP is collaborating with the Helix Nebula Science Cloud project to ensure that when the project enters its pilot phase in 2018, users from the ELIXIR community will be able to use the ELIXIR AAI to access these cloud resources using an allocation purchased by EMBL as a partner in the 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 EGI's accounting model 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).

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.

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[10] which has helped to define the scope and direction of the work. The vision is being refined within the community through the EOSCpilot project[11] and will be implemented through the EOSC-Hub project due to start in January 2018.

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.

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).

---

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

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.

Appendix B: List of Technical Use Cases

More details on the TUCs and the analysis process that derived them can be found in the previous version of the ELIXIR Technical Services Roadmap document (see reference 3).

Table 8. List of the identified Technical Use Cases

ID	Use Case	Description	Success Criteria
1	Federated ID	Organisations expose through Identity Provider services the means for individuals to identify themselves with different levels of assurance to the ELIXIR identity.	A user can use their institutional credential to gain access to their ELIXIR Identity.
2	Other ID	Use of internet identities (e.g. Google, LinkedIn, ORCID, …) with different levels of assurance to the ELIXIR identity.	A user can use their internet identity to gain access to their ELIXIR Identity.
3	ELIXIR Identity	The ELIXIR identity is used as the basis for accessing ELIXIR services.	A user can access and review their ELIXIR ID and use this ID to gain access to their ELIXIR services.
4	Cloud IaaS Services	Bring together centrally the information needed for a user (researcher or analysis service provider) to gain access to a national or regional ELIXIR Cloud service	A user can go through the process of discovering a cloud service (national or regional) before applying and reading the documentation to use a cloud service.
5	HTC/HPC Cluster	Bring together centrally the information needed for a user (researcher or analysis service provider) to gain access to an ELIXIR High Throughput Computing or High Performance Computing (non-PRACE) service.	A user can go through the process of applying and reading the documentation to use a HTC/HPC Cluster service.
6	PRACE Cluster	Link other ELIXIR services (e.g. data, services) with PRACE High Performance Computing systems so a user (researcher or analysis service provider) can access an approved allocation.	A user can apply for access to a PRACE system through their processes.
7	Network File Storage	Support for anonymous read-only, authenticated read-only or authenticated read-write file access to network storage by a user (researcher or analysis service provider).	Network accessible non-local storage space where a user can retrieve or store a file.
8	File Transfer	Schedule the movement of a file by a user (researcher or analysis service provider)	A user can move a file from A to B

		from authentication locations A to B by command line, web service or web page. Depends on TUC 3.	
9	Infrastructure Service Directory	A list of the infrastructure services that might be offered by a site including technical and contact details that is human readable and machine accessible.	A service provider can register their details and their service. A
10	Credential Translation	A tool that converts the ELIXIR ID into a credential that can be used to access a particular service. For instance a federated identity could be converted into a short-term grid proxy.	A user tries to access a service with their ELIXIR ID which is converted on demand to another credential that is accepted.
11	Service Access Management (previously Group/Attribute Management)	A web accessible tool that manages groups and their membership. It allows users to request or withdraw membership of a group. Should also support a user discovering groups and requesting membership and being invited to join.	A user should be able to create a group and add/remove members from the group, in addition to managing their own group memberships in order to gain joint access to services.
12	Virtual Machine Library	A source of virtual machine images with different levels of endorsement and potentially group access management that can be selected for running on cloud services.	A user can select a virtual machine image for running on their cloud. An application developer can upload an image for use by others.
13	Container Library	A source of containers of common software components that can be deployed locally as needed.	A user can upload a container into the library. When required, the container is deployed to provide the required application environment.
14	Module Library	A library of modules of common software components that can be deployed locally as needed.	A user can upload a module into the library. When required, the module is deployed and enabled on a HPC service to provide the required application environment.
15	Data Set Replication	System that will utilise the File Transfer protocol to replicate a Data Set from between major centres given the announcement of a data set release and prior configuration as to whom the data set should be replicated to.	A data provider should be able to move a complete data set from A to B with the system dealing with failures and retries.

16	Infrastructure Service Registry	A registry of currently available and accessible infrastructure services. This dynamic list should match the offered in the Infrastructure Service Directory (TUC 9) however some of these services may not be available, so TUC 16 reflects the current reality.	A service should regularly update its status in the registry with its current status information.
17	Endorsed Personal Data or Compute Access Management	An entitlement to access a specific service may need additional process (e.g. scientific review, phone number verification) that requires additional out of band activities. The authorised individuals have access to the service.	The entitlement manager must be able to configure the workflow and manage the requests and the stages of the authorisation process.
18	Cloud Storage	Storage that can be allocated and attached to virtual machines running in cloud services.	The user can specify a storage device holding their data that can be used by a running virtual machine.
19	PID and Metadata Registry	A service that links a PID (Persistent Identifier to metadata relating to a data file/set. The same data file/set may be registered with multiple physical locations under the same PID.	A user is able to obtain a PID and register their data sets (with location and metadata) and to search the metadata to find data sets.
20	Federated Cloud IaaS	A consistent standards based mechanism is available where through one application a user can gain access to multiple cooperating services.	A user applies once and then by changing the endpoint are able to use any cooperating service.
21	Operational Integration	The services (AAI, cloud, storage, etc.) being offered as part of ELIXIR are monitored to ensure their consistent availability for others within the collaboration.	A user sees that the services they depend upon are reliable. Service providers are able to ensure their services are accessible.
22	Resource Accounting	A portal that is able to view the consumption of services (e.g. CPU time, service invocations, storage, data sets) by individual users, by projects/groups across different services, or provided by ELIXIR nodes. Monitor what services & data sets the users access.	A user is able to see what they have consumed across different services (e.g. cloud, storage) across different projects. A service provider is able to see which users and which projects have consumed their service. A project is able to see the usage that has taken place by user, service and site.

23	Federated HTC/HPC Cluster	A consistent mechanism is available where through one application a user can gain access to multiple cooperating services.	A user applies once and then by changing the endpoint are able to use any cooperating service.

Appendix C: ELIXIR-EXCELERATE WP4 PY2 Management Report

The management report for PY1 can be found in Appendix C of D4.1.

Task 4.1 Leadership

Work package 4 is led by ELIXIR FI, CZ and EMBL-EBI nodes through an Executive Committee (ExCo) that meets at least once a month informally in person and by scheduled teleconferences. The ExCo is supported by a Platform Coordinator, located at the ELIXIR Hub, who ensures that activities of the platform are in line with the activities of the other ELIXIR platforms (Tools, Interoperability, Training), of the ELIXIR Use cases and the ELIXIR work programme. PY2 has started to consolidate the startup activity of PY1 by focusing ECP activities on:

- integrating services to meet the needs of the driving scientific use cases and training through the established Technical Use Cases.
- building on the focused support provided in PY1 for the MMG use case with effort in PY2 on the Human Data use cases.
- continuing to build partnerships with the e-Infrastructure community and other service providers.

Subtask 4.1.1: Management and Coordination

Highlights
- Establishing the ECP 'brand' by attending meetings and presentations.
- Internal WP4 F2F meetings in Stockholm (February) and Prague (June).

September 2016 saw the completion of PY1 and a webinar relating to the analysis of Marine Metagenomics data using the AAI and cloud resources coordinated by WP4[12]. It was decided that in PY2 there would be a focus on Human Data and in September 2017 a workshop was held with WP9 to plan the joint activities for PY2. This work was reviewed at the F2F meetings in February and in June, with regular catch up meetings in between. Progress tracking of this and other work within WP4 was limited by the difficulty in the ELIXIR-Hub filling a full-time project management role for the ECP. Someone was in position briefly between September and December 2016. which enabled meetings between all the task leads to be facilitated every 2-3 weeks. These task level meetings were started again in June 2017 with the appointment of an ECP coordinator.

---

[12]

https://www.elixir-europe.org/documents/elixir-webinar-elixir-compute-platform-roapmap-november-2016

The maturity of the Technical Use Cases (TUCs) introduced in D4.1 is updated in the following sections, can be classified as:
- Proof of Concept (PoC) Services
- Emerging Services (Prototype)
- Mature Services (Production)
- Legacy Services

More details of these classifications can be found in D4.1 Appendix C.

Subtask 4.1.2: Provide a gateway to use European e-Infrastructure services for ELIXIR

Highlights
- Integration of the ELIXIR AAI into the EGI AAI Gateway allow ELIXIR to access EGI's Services
- Development of the Reference Data Set Distribution Tool with the support of EUDAT
- Review of EGI's Service Level Agreements and Operational Policies for adoption by the ECP
- Integration of the ELIXIR AAI into the EUDAT B2ACCESS (prototype)

European e-Infrastructures (e.g. GÉANT, EGI, EUDAT, PRACE) and emerging during PY2 EOSC, provide services that could potentially be of benefit to the ELIXIR Compute Platform. However, one of the challenges of any technical integration is identifying the required services and establishing the appropriate technical and human integration.

Collaboration with EGI has continued with the integration of ELIXIR AAI into the EGI AAI gateway, which has enabled researchers to use the ELIXIR AAI to gain access to EGI services such as GoCDB, AppDB, the Helpdesk and the Operations Portal.

Development of the Reference Data Set Distribution Service with the support of EUDAT2020 project has continued. A prototype of the service is due to be released in September/October 2017. The deployment, wider scale testing and uptake of the service will be supported by an Implementation Study (funded by the ELIXIR-Hub) that will run from September 2017 through to March 2018.

In PY2 the future e-Infrastructure landscape changed significantly with the introduction of the European Open Science Cloud (EOSC). In January 2017, the EOSCpilot project started with representation from the ELIXIR-Hub and EMBL-EBI. Through open calls for Science Demonstrators, the EOSCpilot will be supporting the work of PanCancer, human data analysis and structural biology. The EOSC-Hub project will start in January 2018 and will include an ELIXIR Competency Centre (following on from the work undertaken in the EGI-Engage project) which will focus on demonstrating how ELIXIR Core Data Resources can be packaged for deployment on other sites alongside data provided through the Reference Data Set Distribution Service.

Helix Nebula Science Cloud project is using a pre-commercial model to procure IaaS cloud resources. EMBL is one of the partners participating on behalf of the ELIXIR community. In 2018 some cloud resources will be made available through EMBL to the ELIXIR community using the ELIXIR AAI as the access mechanism. Work in PY2 has seen the commercial cloud providers integrating with the ELIXIR AAI to undertake this work.

Table 10. TUCs supporting the use European e-infrastructure services for ELIXIR

TUC ID	Status PY2 (PY1)	Comment
9 - Infrastructure Service Directory	Emerging (PoC)	Preliminary integration completed in PY1 has been migrated from test instances to the production services in PY2. Work continues to review the EGI SLAs and various operational and security policies to understand their applicability and acceptability to ELIXIR.
21 - Operational Integration	Emerging (PoC)	The integration of ELIXIR sites into EGI's operational infrastructure continues.
22 - Resource Accounting	Emerging (PoC)	Work continues to establish the ELIXIR related usage at sites running the EGI Federated Cloud.

Subtask 4.1.3: ELIXIR technical community building and knowledge exchange

Highlight
- ELIXIR Compute Platform expert colleague network was established
- EGI and GÉANT e-Infrastructures have provided dedicated contact points for ELIXIR

The WP4 community participated in the ELIXIR All-Hands in Rome in March 2017[13] and face to face technical meetings with WP4 and contact points from the use cases in February and June 2017. The ECP coordinator meets weekly with the other ELIXIR Technical coordinators (Interoperability, Tools and Data) to align objectives and actions. The ECP is now being invited as a partner in a number of IS being led by other ELIXIR Platforms.

---

[13] https://www.elixir-europe.org/events/elixir-all-hands-2017

Task 4.2: User Facing Support

This task establishes the user facing support mechanisms to the ELIXIR Compute Platform. Its main consumers are the ELIXIR-EXCELERATE Use Cases (WP6 to 9) and other ELIXIR activities (e.g. ELIXIR-EXCELERATE WPs, ELIXIR pilots, and external projects like EC funded Centres of Excellence or Virtual Research Environments).

Subtask 4.2.1: Technical requirements

Highlights
- TUCs have been updated in PY2 (see Appendix B) as the understanding of the scientific and training use cases have evolved.

This task will collectively maintain and evolve the requirements for the ECP from the scientific and training use cases and prioritise their delivery through the technical service available within the ELIXIR Nodes by capturing these requirements as Technical Use Cases (TUCs). The latest mapping is shown in Appendix B.

Subtask 4.2.2: User support and integration

Highlight
- Planning the expansion of the support helpdesk from beyond a simple contact point to include ECP services.

The development of the support function for the ECP has continued in PY2. The fact that the technical expertise lies within the nodes will be taken into account. Current plans foresee a virtual central helpdesk function that can provide 1st line support where issues will either be resolved by the helpdesk operators or passed on to the relevant distributed 2nd line support units. The 2nd line support units can be varied: being able to provide information (e.g. who should we contact about collaborating in clouds?), able to resolve specific technical issues (e.g. my ELIXIR credentials will not allow me to login to the EMBL-EBI Embassy Cloud), to more broader technical issues (e.g. how can I get my workflow to run on the ECP?). If not resolvable by the second line support unit, third line support units may be available.

A basic contact email alias (contact-compute@ELIXIR-europe.org) has been in operation during PY2.

Work has now focused for PY3 on defining a helpdesk for WP4 activities. This work will start with the ELIXIR AAI support service and will consider integrating support for ELIXIR cloud service providers later in PY3. For each function a clear service

definition, documentation, and contact points are needed as well as a defined operating procedure as to how the assignment, tracking and resolution of tickets is overseen and assured. For the currently listed cloud resources there will need to be a clear model as to how access to these cloud resources will be granted. This support model could be made available to other platforms within ELIXIR and could also provide a contact route for queries to be directed to individual nodes.

Task 4.3 Technical Infrastructure Integration

Task 4.3 integrates existing software and e-Infrastructure solutions to support the implementation of the ECP. The resourcing levels within WP4 have required a strategy of re-use and adaption of services rather than new implementation activities. The role of the individual Node experts working in WP4 is to drive the practical integration of existing Cloud, Compute and Data Services either from the ELIXIR Nodes, or from National and European e-Infrastructure by testing, contributing to documentation and integration with small-scale programming and scripting where needed.

This technical infrastructure integration task is therefore at the heart of the ELIXIR Compute Platform. In order for distributed services to be easily utilised by a distributed life-science community, it is necessary to achieve some coherence on the services, interfaces and operational practices that are being used, so users do not need to re-write their applications every time they move to use compute/cloud/storage services at a different site.

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

Highlights
- The key components, including the ELIXIR Proxy IdP, have entered a mature status
- Developed proof-of-concept services to serve the needs of the human data use case, including multi-factor authentication and bona fide researcher management service (funded by an implementation study).
- Organised training[14] for Relying Services funded by the AARC2 project.
- Agreed on the requirements specification[15] for the Life Science AAI together with the CORBEL WP5 participants (funded by the AARC project).

The ELIXIR AAI has been driven primarily from the ELIXIR FI and ELIXIR CZ nodes with strong support and feedback from across the work package 4. Successful

---

[14] https://docs.google.com/document/d/1-qI5li_vpcrxn6nkwWZn3eMgIlf2qPSwD9k9rfJnXls/edit#
[15]
https://docs.google.com/document/d/1Rcl6sTqyBEnhsxlZdGKUeWZJQ3F11QTBeQOPNKwo8JQ/edit

collaborations established with the AARC project in PY1 has led to ELIXIR participating in the AARC2 project in PY2.

The credential translation services being explored through the AARC project have been used by the ELIXIR AAI to provide X.509 certificates and proxy certificates that can be used in other ELIXIR Infrastructure Services.
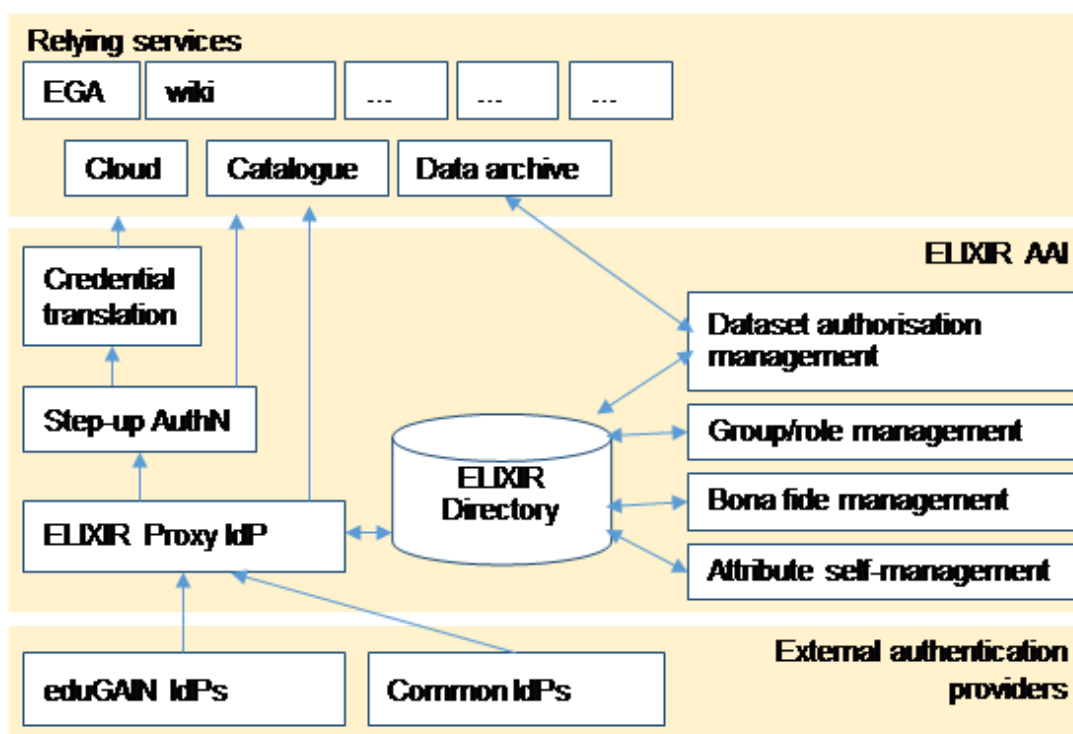

Figure **11.** High-level picture of the ELIXIR AAI design.

Table 11. TUCs supporting authentication and authorization access

TUC ID	Status PY2 (PY1)	Comment
1 - Federated ID	Mature (Emerging)	Currently the ELIXIR AAI delivers only one (low) assurance level due to the inability of the Identity Providers to indicate the assurance level they can support.
2 - Other ID	Mature (Emerging)	Supported internet identity providers: Google, ORCID, LinkedIn.
3 - ELIXIR Identity	Mature (Emerging)	The basic ELIXIR identity has been established through the ELIXIR AAI and is now being used by relying services (e.g. ELIXIR Intranet) to drive authentication and access control decisions. A 'two-factor' authentication demonstrator has been built and discussions continue on how this will be integrated with the ELIXIR AAI.

10 - Credential Translation	PoC (PoC)	The prototype is based on the CILogon software and RCAuth.eu service managed by NIKHEF. It allows users to generate X.509 certificates and proxy certificates to access 'grid' software such as GridFTP, and also provides a OpenID Connect interface that is being integrated into the Globus Transfer service.
11 - Service Access Management	Mature (Mature)	Based on the Perun software from ELIXIR CZ it provides a web based interface for authorising and managing access to different virtual organisations and groups within ELIXIR.
17 - Endorsed Personal Data or Compute Access Management	PoC (PoC)	A PoC has been demonstrated by EGA in WP9.3.2 where a VM uses EGA's APIs to manage and restrict access to sensitive data within a cloud environment.

Subtask 4.3.2: Cloud and Compute integration

Highlights
- Deployment of EGI Federated Cloud environment at EMBL-EBI.
- Feedback to EGI on how their cloud deployment process could be improved which has led to a simpler appliance based deployment method.
- Several nodes have or launched services that can potentially be integrated to ELIXIR Cloud & Compute.

PY2 has seen the continued emergence of cloud service providers affiliated to ELIXIR Nodes into the ECP. Integration into the EGI Federated Cloud is ongoing awaiting the full adoption of the ELIXIR AAI through the EGI AAI gateway across all the EGI Operational Tools and final review of EGI's policies relating to their operational infrastructure.

Cloud service providers at ELIXIR-FI, ELIXIR-CZ and EMBL-EBI have supported the scientific use cases relating to Marine Metagenomics and Human Data. Elements of the two marine metagenomics workloads can be deployed through the EGI Federated Cloud interface. It is expected that the EGI operational model will be the initial operational model being used within the EOSC activities. The human data cloud activities have been working primarily with ELIXIR-FI and are now planning to deploy a similar environment at EMBL-EBI.

Table 12. TUCs supporting Cloud and Compute integration

TUC ID	Status PY2 (PY1)	Comment
4 - Cloud IaaS	Complete	Available information on the cloud services within ELIXIR

Services	(Complete)	has been collected from the ELIXIR Compute Platform partners[16] and will be made available on the ELIXIR website. This information is now being supplemented by a more detailed capacity survey.
5 - HTC/HPC Cluster	Later	Currently, remote access to HPC or HTC services has not been identified as a priority from the Scientific Use Cases.
6 - PRACE Cluster	Later	Access to PRACE class computing services has not been identified as a priority from the Scientific Use Cases.
12 - Virtual Machine Library	(Emerging) PoC	Some early adoption has taken place within the PhenoMeNal and BioExcel projects.
13 - Container Library	PoC (PoC)	The BioShadock[17] service is provided by the French node but is not currently integrated with the ELIXIR AAI.
14 - Module Library	Later	Given the reduced priority for cluster computing (TUCs 5 & 6) from the scientific use cases, there is a corresponding low priority for a Module library that would configure such cluster environments.
18 - Cloud Storage	Later	There are a number of mature cloud storage solutions available to the ELIXIR Compute Platform. As the requirements from the Scientific Use Cases become clearer, the available technologies will be evaluated and a suitable technology deployed.
20 - Federated Cloud IaaS	(Emerging) PoC	While the technology around the EGI Federated Clouds is relatively mature its adoption into ELIXIR is still relatively new and its exact role within ELIXIR, and now EOSC needs to be fully understood.
23 - Federated HTC/HPC Clusters	Later	Federation of HPC/HTC clusters has not been identified as a priority from the Scientific Use Cases.

Subtask 4.3.3: Storage and data transfer

Highlights
- Integration of 9 GridFTP servers at ELIXIR sites into the 'heartbeat' service.

---

[16] https://docs.google.com/document/d/1Dsb_O_EPYrof8WmVtNhBWc5aCnOqjkimdAIbSxYrB0k/edit#

[17] https://docker-ui.genouest.org/app/#/

A mesh of file transfer endpoints has been established, which is being used for 'heartbeat' transfers to test the performance and robustness of various data transfer technologies. These file transfer activities have built on top of the ELIXIR AAI service and the Credential Translation (TUC10) service. Furthermore, documentation and deployment scripts for setting up transfer endpoints have been produced and placed on the ELIXIR intranet (https://www.elixir-europe.org/node/1936/documents) and the most recent stats can be seen online[18].

Development of the Reference Data Set Distribution Service (RDSDS) with the support of EUDAT2020 project continues with an initial release planned for early in PY3. The early deployment activity will be supported through an ELIXIR funded Implementation Study running early in PY3.

Work on deploying an instance of the FTS3 file transfer service developed within the High Energy Physics for evaluation by the ECP continues.

Table 12. TUCs supporting storage and data transfer

TUC ID	Status PY2 (PY1)	Comment
7 - Network File Storage	Later	This TUC needs more information from the scientific use cases. In the ELIXIR Technical Services Roadmap document, all the use cases state the need for temporary storage.
8 - File Transfer	Emerging (PoC)	A procedure for new ELIXIR nodes to deploy GridFTP and to integrate these instances into the heartbeat has been established.
15 - Data Set Replication	PoC (PoC)	Development work continues will plans for a service to enter production in 2018.
19 - PID and Metadata Registry	PoC (Later)	This TUC is dependent on the data set metadata and will be coordinated with WP5.

Subtask 4.3.4: Service Registry

The Infrastructure Service Registry (in subtask 4.3.4) will be used to provide a live picture of the Infrastructure Services that are available in the ELIXIR Compute Platform. The services advertised in the Infrastructure Services Registry are composed of both static information (e.g. contact URL, physical capacity) and dynamic information (e.g. free CPUs, free storage) which may be used by applications to dynamically select

[18] http://sis.ebi.ac.uk/gridftp/ - website can take ~15s to update.

specific services. Such an Infrastructure Service Registry needs to be able to cope with a large number of service updates and a large number of complex client queries.

Some prototyping work has continued in PY2 using the Consul[19] platform.

Table 13. TUCs supporting the development of the service registry

TUC ID	Status PY2 (PY1)	Comment
16 - Infrastructure Service Registry	Later	As service discovery remains a low priority within the ECP, work on this TUC has remained a low priority.

---

[19] https://www.consul.io/