EXCELERATE Deliverable D4.1

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:	A Technical Services Roadmap for supporting Life Science Research in Europe	
WP No.	4	
Lead Beneficiary:		
WP Title	Technical Services	
Contractual delivery date:	31 July 2016	
Actual delivery date:	31 August 2016	
WP leader:	Tommi Nyrönen (CSC), Luděk Matyska (MU)	20: CSC
Partner(s) contributing to this deliverable:	ELIXIR partners	

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), Tony Wildish (EMBL-EBI)

Acknowledgements

Mikael Borg (BILS), Niklas Blomberg (ELIXIR), Rafael Jimenez (ELIXIR), Stephanie Suhr (ELIXIR)

Table of contents

1. Executive Summary

This version of the ELIXIR Technical Services Roadmap (ELIXIR-EXCELERATE deliverable D4.1) has been written at the end of PY1 and represents an update to the previous version that was established prior to the start of the project in Summer 2015. It will be updated again in Summer 2017 following the 2nd Annual meeting of the ELIXIR Compute Platform. The ELIXIR Technical Services Roadmap is a living document that is publicly accessible and commentable and which provides advice as to the current and future implementation activities. As such is subject to change between versions.

This version contains information relating to the technical work undertaken within the ELIXIR Compute Platform during PY1, 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 in the marine metagenomics, forest and crop plants, rare disease and human data communities as well as training provision. Overall, the ELIXIR Compute Platform will need to support these TUCs either 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 PY1 has been extensive with a range of services now becoming available for end-users. The ELIXIR AAI has established a set of services that allows users to create an ELIXIR identity based on a pre-existing identity (e.g. Google or ORCID both 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). This organisational information has then been exposed through the ELIXIR IdP to various relying parties. These include the ELIXIR Intranet, other ELIXIR Services, and the EGI AAI Gateway.

Strong collaborations have been established with External Service Providers allowing us to build upon previous work rather than implementing services ourselves.

The EGI AAI Gateway currently allows ELIXIR Identities to gain access to EGI Tools such as GoCDB (for managing sites and the Infrastructure Services that they provide) and the EGI Applications Database (that allows Virtual Machine images to be contributed and shared within a collaboration).

The AARC project (which includes GÉANT and other relevant partners) works with the ELIXIR AAI to create training materials targeted for providers of scientific services, instructing them on how to adopt federated identity management technologies and processes. The first training event with the AARC project was successfully delivered in early 2016 using the ELIXIR training portal.

The ELIXIR Services that will form the foundation of the ELIXIR Compute Platform are being identified and deployed. Moving files (i.e. data) between sites is a key capability for the ELIXIR Compute Platform. GridFTP servers have been deployed and integrated with the ELIXIR AAI through a credential translation service that converts ELIXIR Identities into X.509 certificates that can identify individuals for site-to-site file transfers. An early prototype of a Data Set Distribution Service that is being developed in collaboration with the EUDAT 2020 project has been used to undertake 'heartbeat' data transfers between sites to assess and help to improve the available network performance.

The availability of individual cloud services within ELIXIR Nodes has been documented. The federation of some of these individual cloud services using the EGI Federated Cloud model is also being explored so that researchers can have a consistent user experience in using ELIXIR clouds regardless of the underlying provider or their technology. Through the EGI-Engage project and the ELIXIR Competency Centre, feedback from ELIXIR has helped improve and simplify the deployment experience.

Technical discussions have taken place with all user communities involved in ELIXIR-EXCELERATE (the four scientific use cases and the training activities) although a pragmatic decision was made in January 2016 to focus initially on non-human data in order to simplify some of the initial work. The main interactions have taken place around the marine metagenomics and the training use cases, with some preliminary discussions around how some of the human data use cases could be supported. In PY2 there will be an increased focus on supporting the secure transport and processing of human data that will start with a joint workshop between WP4 and WP9 to identify some concrete implementations of the current TUCs. To support this work, one of the goals in PY2 is to provide mechanisms (e.g. step-up authentication) to increase trust towards the ELIXIR identity to achieve a level of assurance that satisfies the needs of the scientific service providers.

Service access processes and the level of federation required for cloud and compute services residing in the ELIXIR Nodes will be a focus in PY2. ELIXIR will need solutions for monitoring usage of distributed compute services to support biological data services and data processing. The EGI experience concerning how this can be reliably achieved is valuable.

Although drawing heavily on the funding and the structures provided by the ELIXIR-EXCELERATE project, the ELIXIR Compute Platform will need to persist after the end of the project. To achieve this sustainability it is already drawing on additional services and expertise located within the ELIXIR Nodes.

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

The deliverable is delayed by one month due to the availability of key personnel during the summer period and in the interest of submitting a high quality, complete report.

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	4	Start date or starting event:	month 1
Work package title	**ELIXIR Technical Services**		
Lead	Tommi Nyrönen (CSC – FI) & Luděk Matyska (MU – CZ)		
Participant number and person months per participant 1 – EMBL (34 PM), 6 – NBIC (2 PM; SARA 13 PM, RUG 2 PM), 14 – UPF (23.5 PM), 20 – CSC (48 PM), 23 – UiT (2 PM), 25 – SIB (2 PM), 26 – CNRS (12 PM), 29 – IP (6 PM), 31 – LIU (12 PM), 35 – MU (40 PM), 36 – CESNET (24 PM), 38 – DTU (6 PM)			
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

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.

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.

Deliverables

D4.1 ELIXIR Technical Services Roadmap [11]

ELIXIR Technical Services Roadmap: Incorporates the technical requirements coming from the Use Cases,

details of the service offerings coming from the ELIXIR Nodes and the public e-Infrastructures, and provides a roadmap detailing when new services will appear within the ELIXIR Technical Services registry and the work being undertaken by the service providers to meet ELIXIR's requirements (Task 4.1, task 4.2, task 4.3).

D4.2 Updated ELIXIR Technical Services Roadmap [23]

Updated ELIXIR Technical Services Roadmap: Incorporates the technical requirements coming from the Use Cases, details of the service offerings coming from the ELIXIR Nodes and the public e-Infrastructures, and provides a roadmap detailing when new services will appear within the ELIXIR Technical Services portfolio and the work being undertaken by the service providers to meet ELIXIR's requirements (Task 4.1, task 4.2, task 4.3).

D4.3 ELIXIR Technical Services Roadmap [35]

ELIXIR Technical Services Roadmap: Incorporates the technical requirements coming from the Use Cases, details of the service offerings coming from the ELIXIR Nodes and the public e-Infrastructures, and provides a roadmap detailing when new services will appear within the ELIXIR Technical Services portfolio and the work being undertaken by the service providers to meet ELIXIR's requirements (Task 4.1, task 4.2, task 4.3).

D4.4 ELIXIR Technical Services document [46]

ELIXIR Technical Services document: Describes the ELIXIR Technical Services being offered through the cooperating service providers (ELIXIR Nodes and public e-infrastructures) (Task 4.1, task 4.2, task 4.3).

Annex 1: A Technical Services Roadmap for supporting Life Science Research in Europe

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

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), Tony Wildish (EMBL-EBI)

Acknowledgements

Mikael Borg (BILS), Niklas Blomberg (ELIXIR), Rafael Jimenez (ELIXIR), Stephanie Suhr (ELIXIR)

Key ELIXIR Compute Platform achievements during project year 1

Initial requirements gathered from the use cases (WP6-9 & WP11) to define an ELIXIR Technical Services Roadmap to facilitate the building of the ELIXIR Compute Platform

- Establishing the ELIXIR Identity through the ELIXIR AAI
- Deploying GridFTP services and establishing a data transfer 'heartbeat'
- Identifying the currently available cloud service providers within ELIXIR Nodes

Collaboration with the AARC project

- Credential translation services for X.509 certificates based on ELIXIR AAI
- ELIXIR AAI training materials and events

Collaboration with EGI and the EGI-Engage project

- Support for the marine metagenomics use case (WP6) through ELIXIR Competence Center in the EGI-Engage project
- Feedback on the EGI Federated Cloud deployment model which has led to an improved process for ELIXIR and EGI service providers
- Integration of the ELIXIR AAI into various EGI Operational Tools through the EGI AAI gateway
- Adoption of the EGI Operational Tools (i.e. service directory, monitoring and accounting) to manage the Infrastructure Services within the ELIXIR Compute Platform

Collaboration within the ELIXIR-EXCELERATE project

- WP6 to drive virtualisation of marine metagenomics data analysis software environment to support its deployment on cloud
- WP7 to provide support for metadata gathering from distributed sites hosting plant genomics data, followed later by data processing capabilities
- WP8 and 9 to document security and other requirements for transport and of human data, which will be in focus in PY2
- WP11 to support use of clouds for bioinformatics training

Horizon 2020 grant n. 676559

Table of contents

Horizon 2020 grant n. 676559

Tables and Figures

1. Executive Summary

This version of the ELIXIR Technical Services Roadmap (ELIXIR-EXCELERATE deliverable D4.1) has been written at the end of PY1 and represents an update to the previous version that was established prior to the start of the project in Summer 2015. It will be updated again in Summer 2017 following the 2nd Annual meeting of the ELIXIR Compute Platform. The ELIXIR Technical Services Roadmap is a living document that is publicly accessible and commentable and which provides advice as to the current and future implementation activities. As such is subject to change between versions.

This version contains information relating to the technical work undertaken within the ELIXIR Compute Platform during PY1, 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 in the marine metagenomics, forest and crop plants, rare disease and human data communities as well as training provision. Overall, the ELIXIR Compute Platform will need to support these TUCs either 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 PY1 has been extensive with a range of services now becoming available for end-users. The ELIXIR AAI[1] has established a set of services that allows users to create an ELIXIR identity based on a pre-existing identity (e.g. Google or ORCID both 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). This organisational information has then been exposed through the ELIXIR IdP to various relying parties. These include the ELIXIR Intranet, other ELIXIR Services, and the EGI AAI Gateway.

Strong collaborations have been established with External Service Providers allowing us to build upon previous work rather than implementing services ourselves.

The EGI AAI Gateway currently allows ELIXIR Identities to gain access to EGI Tools such as GoCDB (for managing sites and the Infrastructure Services that they provide) and the EGI Applications Database (that allows Virtual Machine images to be contributed and shared within a collaboration).

The AARC project[2] (which includes GÉANT and other relevant partners) works with the ELIXIR AAI to create training materials targeted for providers of scientific services, instructing them on how to adopt federated identity management technologies and processes. The first training event with the AARC project was successfully delivered in early 2016 using the ELIXIR training portal.

The ELIXIR Services that will form the foundation of the ELIXIR Compute Platform are being identified and deployed. Moving files (i.e. data) between sites is a key capability for the ELIXIR Compute Platform. GridFTP servers have been deployed and integrated with the ELIXIR AAI through a credential translation service that converts ELIXIR Identities into X.509 certificates that can identify individuals for site-to-site file transfers. An early prototype of a Data Set Distribution Service that is being developed in collaboration with the EUDAT 2020 project has been used to

[1] https://docs.google.com/document/d/1cJ3mR8lqfZKRMvSFalSmPbqd1OPU-L6YcUFIRnh1rhQ/edit

[2] https://aarc-project.eu/

Horizon 2020 grant n. 676559

undertake ‘heartbeat’ data transfers between sites to assess and help to improve the available network performance.

The availability of individual cloud services within ELIXIR Nodes has been documented. The federation of some of these individual cloud services using the EGI Federated Cloud model is also being explored so that researchers can have a consistent user experience in using ELIXIR clouds regardless of the underlying provider or their technology. Through the EGI-Engage project and the ELIXIR Competency Centre, feedback from ELIXIR has helped improve and simplify the deployment experience.

Technical discussions have taken place with all user communities involved in ELIXIR-EXCELERATE (the four scientific use cases and the training activities) although a pragmatic decision was made in January 2016 to focus initially on non-human data in order to simplify some of the initial work. The main interactions have taken place around the marine metagenomics and the training use cases, with some preliminary discussions around how some of the human data use cases could be supported. In PY2 there will be an increased focus on supporting the secure transport and processing of human data that will start with a joint workshop between WP4 and WP9 to identify some concrete implementations of the current TUCs. To support this work, one of the goals in PY2 is to provide mechanisms (e.g. step-up authentication) to increase trust towards the ELIXIR identity to achieve a level of assurance that satisfies the needs of the scientific service providers.

Service access processes and the level of federation required for cloud and compute services residing in the ELIXIR Nodes will be a focus in PY2. ELIXIR will need solutions for monitoring usage of distributed compute services to support biological data services and data processing. The EGI experience concerning how this can be reliably achieved is valuable.

Although drawing heavily on the funding and the structures provided by the ELIXIR-EXCELERATE project, the ELIXIR Compute Platform will need to persist after the end of the project. To achieve this sustainability it is already drawing on additional services and expertise located within the ELIXIR Nodes.

2. Introduction

This report provides an update to the ELIXIR Technical Services Roadmap generated in Summer 2015 by the ELIXIR Compute Platform before the start of the ELIXIR-EXCELERATE project[3]. Since then it has been used as the 'blueprint' for the technical activities taking place within the ELIXIR Compute Platform and to support the ELIXIR Scientific Programme 2014-2018. Within the project, the efforts are focussed on the marine metagenomics, forest and crop plants, rare disease and human data communities as well as training provision (ELIXIR-EXCELERATE 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 ELIXIR Compute Platform (detailed in Appendix B). The majority of the requirements have been captured and a set of services drawn from European e-Infrastructures and ELIXIR Nodes are now being established to meet these needs.

Now, after a year of activity, the ELIXIR Technical Services Roadmap is being revised and updated to reflect the experiences of the last year and to help inform ELIXIR's researchers and application developers of the ELIXIR Compute Platform services that will be available for them to use in the coming years. For ELIXIR Nodes and associated Infrastructure Service providers, this report identifies the technologies that should be deployed to enable ELIXIR to provide a consistent set of Infrastructure Services.

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 needs of the research community in biological information, and to influence future development priorities of European e-Infrastructures. Essentially, the role of the ELIXIR Compute Platform is to define a minimal 'neck' of an hourglass that ELIXIR Researchers and Application Developers can build upon and which ELIXIR Nodes and other Infrastructure Service providers can deploy and support. This has led to a strategy of collaborating with existing initiatives and organisations rather than developing new services ourselves. Such a strategy requires an upfront investment in building collaborative relationships, resulting in initially slow progress. However, as the common understanding of goals and trust develops between the collaborating organisations and the use of existing technology and expertise is leveraged, the capability of the overall distributed infrastructure improves as the services will have increased maturity and the responsibility for keeping them online is shared.

During PY1, the ELIXIR Compute Platform coordinated the development and delivery of a geographically distributed Authentication & Authorisation Infrastructure (AAI) as well as Cloud & Compute, Storage and File Transfer Services that are provided by the individual ELIXIR Nodes and which will ultimately be discoverable through the ELIXIR Infrastructure Service Registry. Although drawing heavily on the funding and the structures provided by ELIXIR-EXCELERATE, the ELIXIR Compute Platform will persist after the end of the project and is already drawing on additional services and expertise residing in the ELIXIR Nodes to achieve this sustainability.

[3] Initial technical services roadmap, Summer 2015
https://drive.google.com/open?id=0ByWHBDVpEowoTm92N0VEUnV2UXc

The work undertaken during PY1 across all TUCs and the collaborations that have been established with organisations outside the immediate scope of the project are summarised in Section 3 and detailed in Appendix D.

The 'Motivating Use Cases' section describes the scientific and training Use Cases that are being used to drive the development of the ELIXIR Compute Platform. This section has been updated to reflect the improved understanding gained during PY1 of the use cases in how they will use the ELIXIR Compute Platform. For each use case the supporting TUCs have been identified and prioritised by the users with the support of members of the ELIXIR Compute Platform.

In the 'Supporting the Use Cases' section an assessment of the progress made in each TUC has been completed and TUCs have been classified based on their maturity (Proof of Concept, Emerging (or Prototype) Service, Mature Service in production, Legacy Service that is moving out of production). Those TUCs for which no progress has been made in PY1 are identified and they have been prioritised over the remaining project years.

In the final two sections we describe a high-level architecture of the Compute platform infrastructure and our plans for PY2 of ELIXIR-EXCELERATE. To support the reader in reading this report a comprehensive glossary of technical terms is provided in Appendix A.

3. ELIXIR Compute Platform Activities in Project Year 1

During the first year of the ELIXIR-EXCELERATE project, WP4 has established the necessary management, technical support, and technical structures needed to build the ELIXIR Compute Platform. More details are provided in Appendix C.

3.1. Leadership

The management and leadership have been established in the form of an Executive Committee that meets at regular intervals to coordinate activity within the ELIXIR Compute Platform and towards its external actors. The focus of this interaction has been in two directions: towards the European e-Infrastructures (such as EGI, EUDAT and GÉANT) which are expected to provide some of the Infrastructure Services that the ELIXIR Compute Platform will build upon, and establishing a technical community within the ELIXIR Nodes that will be running some of the Infrastructure Services made available to the research community.

3.2. User-Facing Support

User-Facing Support of the consumers of the ELIXIR Compute Platform has also started. Interaction with the scientific Use Cases and the training activities within the ELIXIR-EXCELERATE project has been established and captured through the prioritised TUCs with practical motivating use cases. In PY1 the priority has been on non-human data, but this focus is now being broadened for PY2. As the ELIXIR Compute Platform starts offering services to its users, the support infrastructure needed to provide a good user experience (e.g. helpdesk, documentation, FAQs, etc.) has been considered and is being planned for in PY2. Another aspect of the User-Facing Support is the collection of requirements from the use cases, their analysis and the maintenance of the ELIXIR Technical Services Roadmap throughout the year.

Horizon 2020 grant n. 676559

3.3. Technical Infrastructure Integration

Finally, the technical activities within the ELIXIR Compute Platform—the implementation of the TUCs (which are defined in detail in Appendix B)—has progressed with work in establishing ELIXIR's interlinked AAI, Storage and Data Transfer, Cloud & Compute, and Infrastructure Service Registry activities. For example, a researcher may use the ELIXIR Compute Platform to discover cloud services from the Infrastructure Service Registry and to use their ELIXIR Identity to provision a software analysis environment on an EGI Federated Cloud service. To undertake their analysis they would need to transfer data from the reference datasets held within ELIXIR's Data Services to the selected cloud services.

3.3.1. Authentication and Authorisation Infrastructure

The AAI task builds on the European federated identity fabric that can authenticate an individual and verify if they are a member of a group (or have a particular role within a group). The task has established the concept of an ELIXIR Identity and encompasses the entire ELIXIR research community, including users whose home organization does not provide federated identities. The ELIXIR AAI also provides additional AAI services such as Proxy IdP, eduGAIN IdPs, Common IdPs, guest login, ELIXIR directory, attribute self-management for users, bona-fide researcher management, group/role management, dataset authorisation management and credential translation.

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. Successful collaborations have been established with the AARC project in terms of providing requirements and benefiting from some of their service prototyping.

3.3.2. Cloud & Compute

The Cloud and Compute Task has identified available ELIXIR Node cloud services[4] for inclusion into the ELIXIR Compute Platform and has started promoting information on how services (e.g. IaaS cloud services) can be consumed by the ELIXIR research community. This takes place in the context of science-specific platforms and workflows defined in the EXCELERATE Use Cases in WP6 to 9 and WP11. The Cloud & Compute task supports the software environment that use cases need for their 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).

3.3.3. Storage and Data Transfer

Data transfers are needed across all scientific use cases and various data transport mechanisms have been investigated to organise data transfers between different

[4] The ELIXIR Cloud & Compute task integrates willing providers into the ELIXIR Compute Platform (e.g. EMBL-EBI Embassy Cloud; national level: CSC, SURFsara, MetaCentrum, CERIT-SC, Nordic Secure Cloud; regional level: GÉANT, EGI federated cloud; collaboration with the commercial sector: Helix Nebula, Amazon, Microsoft, Google)

Horizon 2020 grant n. 676559

ELIXIR data centres[5]. Three common use cases drive storage and data transfer activities: (a) Data replication and data submission to or from ELIXIR Data Resources (EXCELERATE WP3); (b) Services to pull relevant datasets from Data Resources or their replicas to cloud or compute services for detailed local analysis; (c) Data location services to manage and discover data replicas within ELIXIR established to decrease network overload for ELIXIR nodes hosting large data sets to deter *ad hoc* data transfer and storage. The managed access integration will be piloted in the human data use case (ELIXIR-EXCELERATE WP9).

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

3.3.4. Infrastructure Service Registry

An ELIXIR Infrastructure Service Registry will be deployed to provide a live picture of the technical capabilities of the ELIXIR Compute Platform. This registry will complement other registries such as the Tools Registry provided by WP1. The information in the Infrastructure Service Registry is composed of both static information (e.g. contact URL, physical capacity) and dynamic capability information (e.g. free CPUs, free storage). Information in the registry can be used by ELIXIR Services and applications to select which specific Infrastructure Service to use. Scalability is one of the key design criteria since 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. The overall progress of the TUCs in PY1 is provided in **Error! Reference source not found.**. EMBL-EBI and ELIXIR CZ lead this work.

[5] The Data and storage task is led by the ELIXIR-SE, EMBL-EBI and ELIXIR-NL nodes. The "heartbeat" file transfer mesh is currently composed of six active nodes (ELIXIR-SE: BILS, ELIXIR-ES: CRG, ELIXIR-FI: FUNET, CSC, ELIXIR-CZ: CESNET, EMBL-EBI), with another two close to integration (ELIXIR-NL: SurfSARA, ELIXIR-FR: GenOuest) and a further two in the pipeline (ELIXIR-NO: NeLS, EMBL-EBI: EGA) plus Indiana University (USA)

Table 1 Assessment of the overall progress on individual TUCs over PY1

ID	Technical Use Case	Status[6]
1	Federated ID	Emerging
2	Other ID	Emerging
3	ELIXIR Identity	Emerging
4	Cloud IaaS Services	Complete
8	File Transfer	PoC
9	Infrastructure Service Directory	PoC
10	Credential Translation	PoC
11	Service Access Management	Mature
12	Virtual Machine Library	PoC
13	Container Library	PoC
15	Data Set Replication	PoC
17	Endorsed Personal Data or Compute Access Management	PoC
20	Federated Cloud IaaS	PoC
21	Operational Integration	PoC
22	Resource Accounting	PoC

Implementation has not yet started for all TUCs. As indicated below, some of the TUCs are expected to be addressed in future project years. For other TUCs (e.g. 7 & 14), although they were identified during the use case analysis, there is little evidence that these are a priority at this stage. The prioritisation of TUCs is always subject to change in response to user feedback.

Table 2 Identified TUCs marked for later implementation

ID	Technical Use Case	Status	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
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]TUC Status: Proof of Concept (PoC), Emerging Services (Prototype), Mature Services (Production), Legacy Services.

Horizon 2020 grant n. 676559

4. Motivating Use Cases for the ELIXIR Compute Platform

This section provides an overview of each research-oriented use case in the marine metagenomics, forest and crop plants, rare disease and human data communities as well as training provision, and describes how these have been broken down into individual TUCs.

4.1. Use Case A: Marine Metagenomics (WP6)

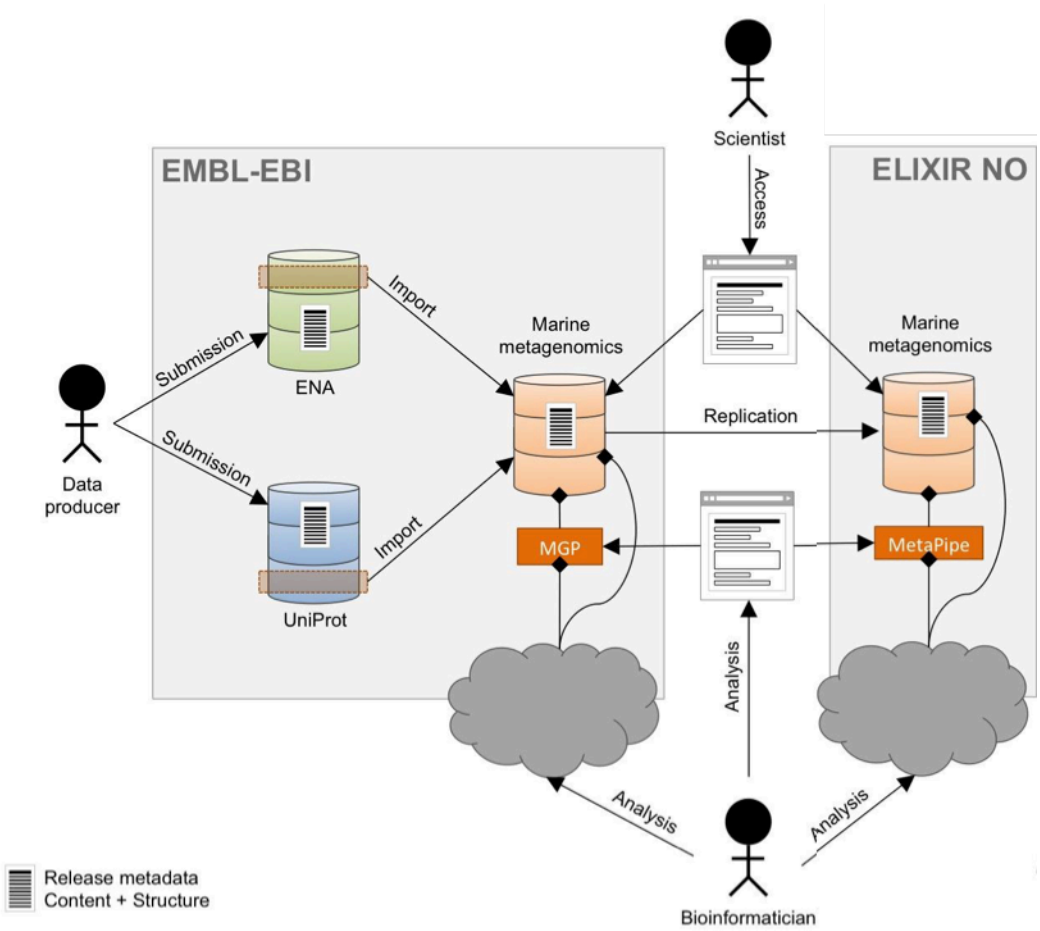

Figure 1 Marine metagenomics use case (EXCELERATE WP6)

Example of Marine Metagenomics use case. The data producer uploads information to different databases, which can feed other databases or even be replicated in along the different ELIXIR Nodes. The scientists can request access to the datasets in the databases. The bioinformaticians can access to datasets and process or analyze them using cloud or computing resources.

The marine metagenomics infrastructure provides a web-based portal that will act as a driver for research and industrial innovation. The portal will be populated by pipelines[7] (including curated Galaxy workflows) running on internal High Throughput

[7] Definition available in Appendix A: Glossary of terms

Horizon 2020 grant n. 676559

Computing (HTC) clusters and using data that is generated from within the collaboration but also collected from other data sources. These pipelines will be ported from the current cluster to the future cloud infrastructure.

Table 3 List of Technical Use Cases (TUCs) supporting the Marine Metagenomics use case

ID	TUC	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. These credentials may need to be translated to other systems (TUC 10).
5	HTC/HPC Cluster	An HPC cluster currently provides the analysis capacity for the pipelines to populate the portal. A Norwegian e-Infrastructure cluster (Nortur) 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. For users to access different Cloud IaaS's consistently, these services will need to be Federated (TUC 20) and integrated into an operational infrastructure (TUC 21) where Infrastructure Services can be declared in an Infrastructure Service 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 and CSC's cPouta. The EGI-Engage ELIXIR Competence Centre has provided support for example to use cPouta. CSC's Ansible playbooks (recipes) to setup a Spark data analysis virtual cluster have been made available.
7	Network File Storage	Needed to provide temporary storage of 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	Needed to move data from primary to temporary storage. Over 1 PB (maximum estimate) is expected to be replicated over 4 years including data from various archives (TUC 15).

4.2. Use Case B: Genomic and Phenotypic Data for Crop and Forest Plants (WP7)

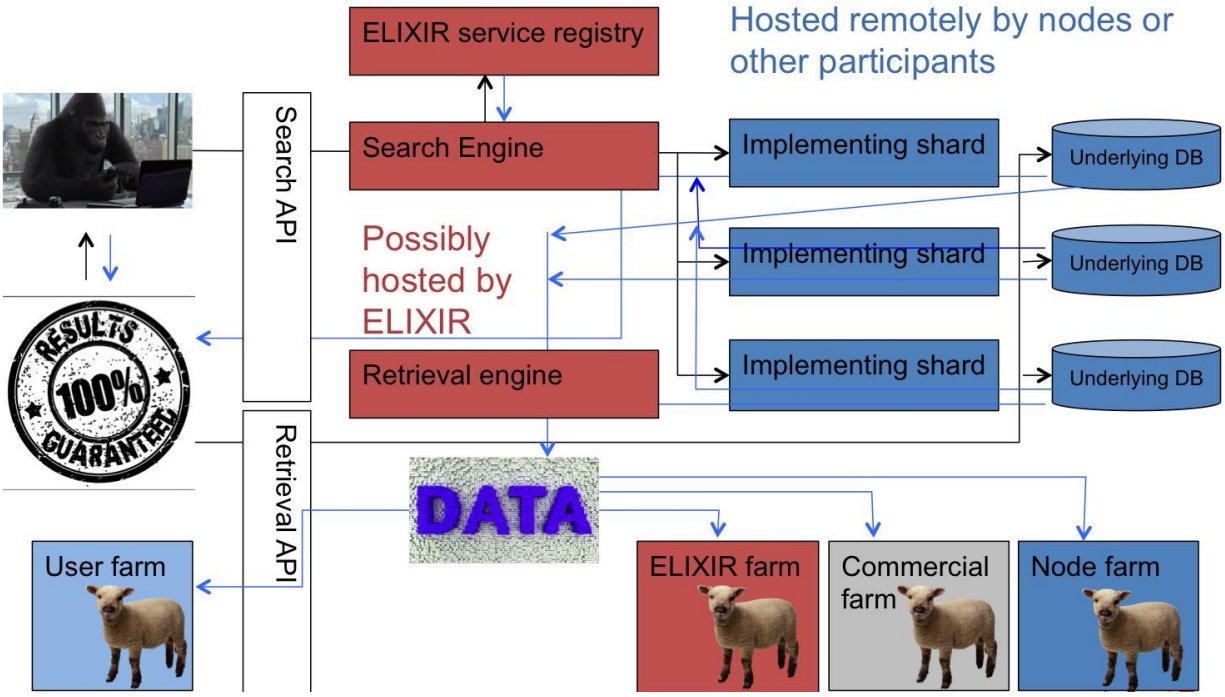

Figure 2 Crop and forest plants use case (EXCELERATE WP7)

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 Service 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.

Horizon 2020 grant n. 676559

- **Analysis of Identified Data**: Moving the identified data to an available compute environment would make use of TUCs 5, 7 and 8. TUC 4 could be used to provision a TUC 5 compute environment.

Table 4 List of Technical Use Cases (TUCs) supporting the Genomic and Phenotypic Data for Crop and forest plants

ID	TUC	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	Provides 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.
16	Infrastructure Service Registry	Registry that holds the location of conformant data discovery instances and 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.
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 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.

4.3. Use Case C: Rare Disease and (WP8) and Use Case D: Human access-controlled Data (WP9)

These two scientific use cases involving the use of human data were seen to have essentially identical technical requirements.

Horizon 2020 grant n. 676559

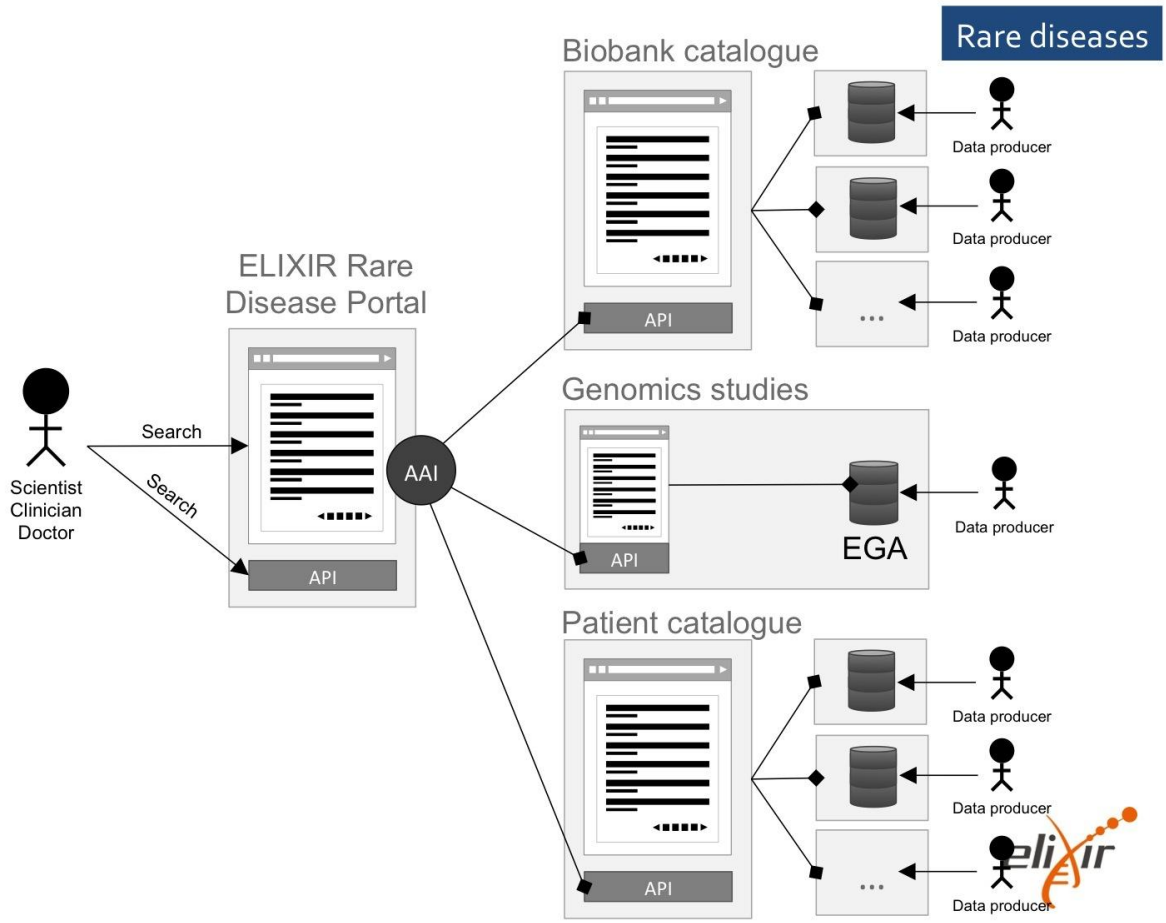

Figure 3 Rare disease use case (EXCELERATE WP8)

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.

The rare disease use case is based on supporting research around rare (1 in 2000 people) chronic or genetic diseases that uses EGA as its data source—access to which is controlled. The metadata concerning individual patients (i.e. their illness, treatments, outcomes), patient samples stored in a biobank, and any sequenced material stored in EGA are 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.

Horizon 2020 grant n. 676559

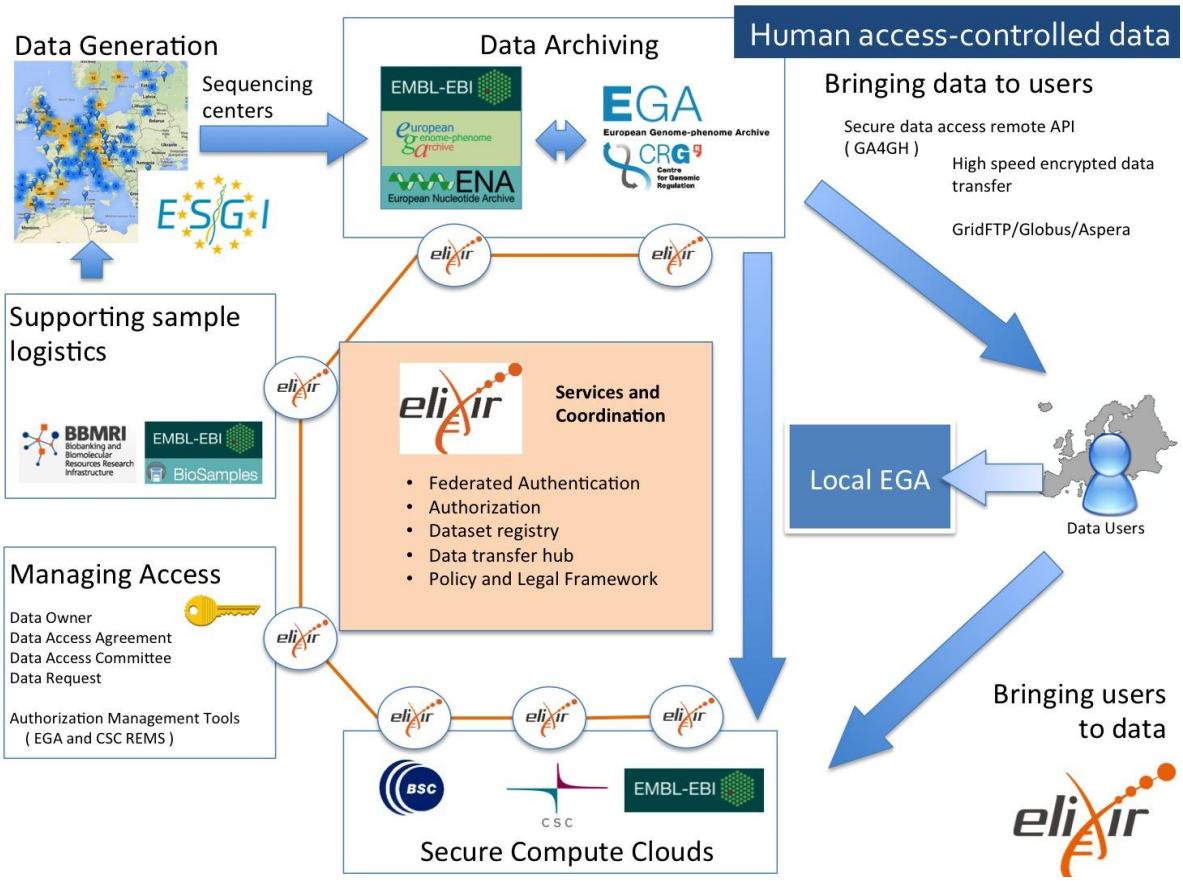

Figure 4 Human data use case (EXCELERATE WP9)

The WP9 use case around human access-controlled data 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).

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 model has been endorsed 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 Service 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.

4.4. Use Case E: Training (WP11)

A more detailed description of the training requirements has been captured elsewhere[8]; results are summarised here. The training use case has prioritised the identity and cloud (IaaS) TUCs (1 to 4).

[8] https://docs.google.com/spreadsheets/d/1fCcbbOpNHg9TGmSHxBDowDsnHN7YHhmq79lj5fPL_Gk

Table 6 List of Technical Use Cases (TUCs) supporting the training use case

ID	TUC	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 the ELIXIR ID is established it will be used in addition to the Federated ID.
2	Other ID	Useful for newcomers to training sessions, without any pre-registration.
3	ELIXIR ID	This is needed to authenticate trainers for accessing the central services, training portals, internal servers etc.
4	Cloud IaaS Services	A cloud infrastructure is needed to run e-learning courses, training courses etc. Expected usage is one virtual machine (VM) for one trainee, preferably on one site.
5	HTC/HPC Cluster	Only if training HTC/HPC are important and needed to perform 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; download VM image from library to training environment.
12	Virtual Machine Library	It is foreseen that the training environment will use VM images, where software used in training is pre-installed (e.g. Galaxy image).
13	Container Library	Similar/alternative for TUC 12, software pre-installed in Docker images is expected to be used.
18	Cloud storage	Useful for sharing data between VMs (referenced data or pre-installed tools in Galaxy/Chipster).
22	Resource Accounting	Expected to be used for user charging and for training resource planning.
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.

4.5. Other ELIXIR Use Cases

The ELIXIR Beacon project is part of an ELIXIR core-funded study with the aim to initiate a GA4GH Beacon[9] API in a number of partnering ELIXIR Nodes by the end of 2016. Initially, the Beacon service provides a simple query interface for genome information. The Compute platform supports the ELIXIR Beacon project by integrating the required authentication and authorization tools with the Beacon prototype implementation. The work started in June 2016 and is expected to continue through PY2.

5. Supporting the Use Cases

The priorities for PY2 were determined based on the TUC prioritisation by the motivating use cases and the work the ELIXIR Compute Platform has undertaken in PY1 to implement the TUCs.

[9] https://genomicsandhealth.org/work-products-demonstration-projects/beacon-project-0

Table 7 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

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

6. High-level Architecture

The ELIXIR Compute Platform's AAI, Cloud, and Storage & File Transfer Tasks have been logically organised using three layers that describe the positioning of the task with respect to their relying and external parties. Each task has described a set of technical capabilities in order to fulfil TUCs that can be produced by the ELIXIR Compute Platform directly or in collaboration with research infrastructures or other third (commercial) parties. When the ELIXIR Compute Platform delivers the TUCs required by a motivating use case, a chain of dependencies is created. The architectural outline helps the ELIXIR Compute Platform to understand and manage its dependencies, and thus ensure quality of the services for the relying parties.

The ELIXIR Compute Platform architectural outline includes:

- **Relying Services** (e.g. ELIXIR intranet, biobank data access portal) are typically scientific service providers that receive technical service capability from the ELIXIR Compute Platform
- **ELIXIR Compute Platform** service components are operated directly by the ELIXIR Nodes and consumed by the Relying Services
- **External Service Providers** allow ELIXIR to leverage investments made nationally, for example by EC projects or commercially by other Service

Providers (e.g. the ELIXIR AAI architecture leverages federated identity management infrastructure and trust management fabrics provided by GÉANT).

Ongoing work of the ELIXIR Compute Platform aims to describe service integration across External Service Providers to the ELIXIR Compute Platform and Relying Services in order to deliver for the ELIXIR-EXCELERATE motivating use cases.

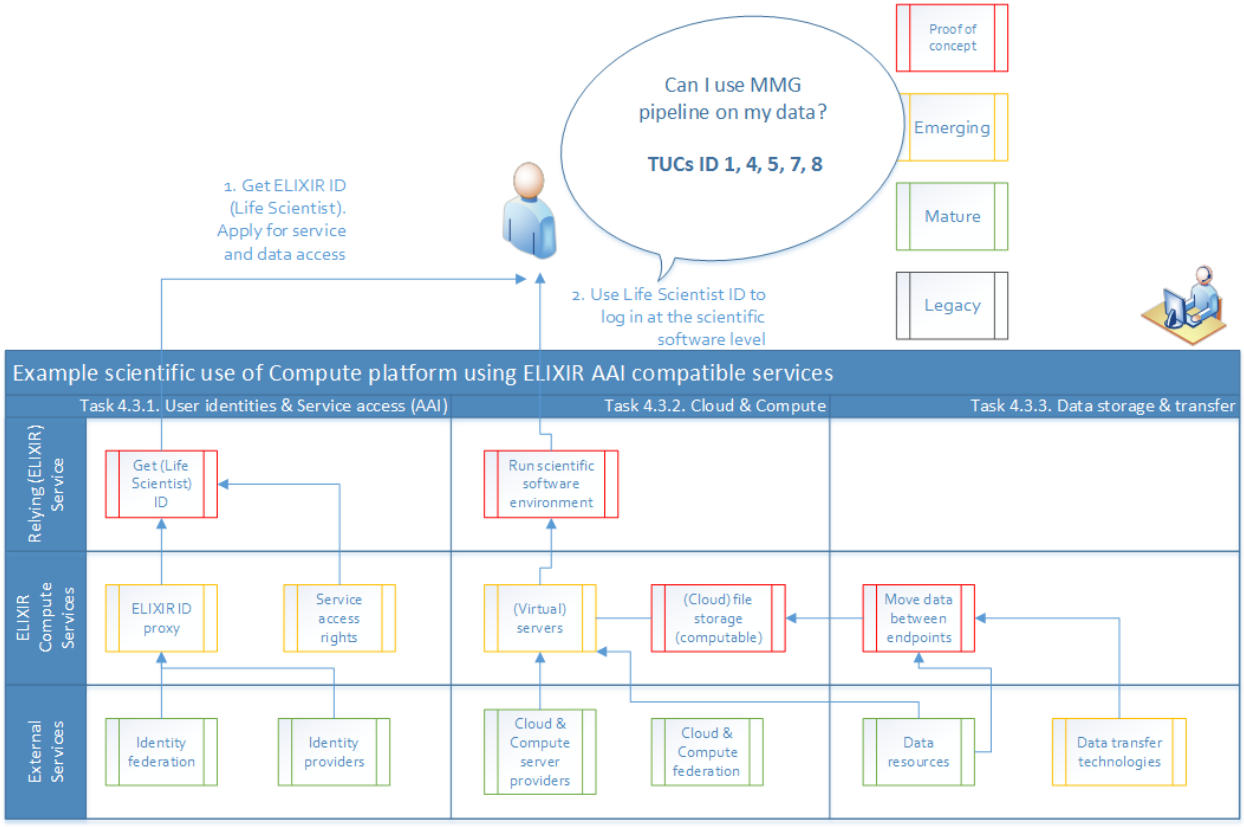

Figure 5 Illustration of dependencies (arrows) from the ELIXIR-EXCELERATE WP6 use case with the various TUCs (i.e. scientific service requests for an ELIXIR Compute Platform service)

WP6 request will be met by the ELIXIR Compute Platform and collaborating External Service providers, both of which will need to be relying parties of the ELIXIR AAI. Each box (red, yellow, green, gray) forming the service process is a process itself, and has dependencies (arrows). For example, cloud & compute provider supplies servers for a bioinformatics service provider running a complex scientific software pipeline service, or a data set is transferred from a user's local computing environment to a computable storage with reference data from an archive, seen by the server(s).

7. Plans for Next Year

The ELIXIR Compute Platform needs to utilise the work of existing service providers and technologies wherever possible, partly to reduce the time and cost of bringing newly developed services into operation, but also because ELIXIR-EXCELERATE and the ELIXIR nodes are limited in the availability of human resources necessary for new technical developments.

Consequently, the strategy of the ELIXIR Compute Platform will remain focussed on collaborating widely within the ELIXIR community and on the services emerging from the Nodes, as well as with European e-Infrastructures projects and other service providers. Collaboration with the AARC project (and AARC2, which was recently

approved for funding by the European Commission) will continue with the intention that the ELIXIR AAI model will become more widely adopted in the life-science community. Following an initial training event on the ELIXIR AAI, delivered by ELIXIR in collaboration with the AARC project, a second, broader ELIXIR AAI training event for biomedical service providers in the context of the CORBEL project[10] is planned. The training material coming out of these events will be refined and made available for independent use via the ELIXIR Training Portal TeSS[11] (EXCELERATE WP11).

An additional goal of the ELIXIR Compute Platform is for the ELIXIR IdP to be accepted by e-Infrastructures as a way to provide user credentials to access services from other European e-Infrastructures. EGI has already adopted this approach and EUDAT have expressed their interest in also using this model. EUDAT are supporting the development of a Reference Data Set Distribution service within the EUDAT 2020 project, and a prototype of this service is expected to be deployed within the ELIXIR Compute Platform during PY2.

PY2 will see an increased focus on providing solutions for research with access-controlled human data. WP4 will hold a workshop with WP9 early in PY2 to increase the interaction and start developing support around use cases for secure archiving, dissemination and analysis of human access-controlled data. This will include the provision of higher assurance levels within the ELIXIR AAI, which will be explored in collaboration with GÉANT and the AARC/AARC2 projects.

In addition, the ELIXIR Compute Platform's federation model will continue to develop around the ELIXIR AAI and other aspects as required by the needs of researchers and service providers. The inclusion of other cloud services (e.g. commercial cloud services coming through GEANT, Helix Nebula and other procurement mechanisms) alongside academic cloud services will need to be understood. This will help illustrate ELIXIR can provide a clear pay-for-use model for SME's that can be used to analyse ELIXIR data using service providers able to support commercial users that is integrated with the ELIXIR AAI.

An initial step has been taken to document and publish the various service access processes, and opportunities for harmonisation will be explored. Alongside service access, ELIXIR will need solutions for monitoring service usage (of compute, cloud and storage) so that virtual access costs can be passed on to the consuming projects (or, ultimately, their funding agencies), and in order to start developing indicators and metrics on the usage of different services within the ELIXIR Compute Platform. The EGI experience in collecting accounting records will provide a valuable starting point in this initiative. A key target for PY2 is to identify the service costs around the supported use cases (EXCELERATE WP6 to 9 and WP11).

The deployment of GridFTP as the primary file transfer mechanism will continue while other approaches are explored. An integration between the ELIXIR AAI and Globus is being explored that would allow members of the ELIXIR AAI to gain access to the Globus file transfer coordination services (already being used by EGA) which would be able to use the deployed GridFTP endpoints. Other such third-party transfer coordination services (e.g. WebFTS) may be explored if resources permit.

As illustrated above, although drawing heavily on the funding and the structures provided by ELIXIR-EXCELERATE, the Compute Platform will persist after the end of the ELIXIR-EXCELERATE project and is already drawing on additional services and expertise residing in the ELIXIR Nodes.

[10] www.corbel-project.eu
[11] http://tess.oerc.ox.ac.uk/

Appendix A: Glossary of 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.

eduGAIN: GEANT'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.

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).

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.

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 that 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).

Appendix B Table 1 List of the identified Technical Use Cases (TUCs)

ID	Use Case	Description	Success Criteria
1	Federated ID	Through Identity Provider services organisations expose the means for individuals to identify themselves to the ELIXIR identity with different levels of assurance.	A user can use their institutional credential to gain access to their ELIXIR Identity.
2	Other ID	Use of internet identities (e.g. Google, Facebook, …) 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 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 on how to use a cloud service.
5	HTC/HPC Cluster	Bring together centrally the information needed for a user to gain access to an ELIXIR HTC or HPC (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 HPC systems.	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.	Network accessible non-local storage space where a user can retrieve or store a file.
8	File Transfer	Schedule the movement of a file from authentication locations A to B by command line, web service or web page. Depends on TUC 3	A user can move a file from A to B
9	Infrastructure Service Directory	A list of the Infrastructure Services including technical and contact details that is human readable and machine accessible offered by a site.	A service provider can register their details and their service.
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 group membership. It allows users to request or withdraw from 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 (VM) 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 VM 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 Infrastructure Services available for use. This dynamic list should match the offered in the Service Directory.	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 VMs running in cloud services.	The user can specify a storage device holding their data that can be used by a running VM.
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 through which, with a single 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 Management Report, PY1

Task 4.1 Leadership

Work package 4 is led by the ELIXIR-FI, ELIXIR-CZ and EMBL-EBI Nodes through an Executive Committee (ExCo) that meets at regular intervals, informally in person and via scheduled phone conferences. During PY1 the focus has been on establishing the technical activities within the Compute Platform and linking these activities with complementary activities within ELIXIR (through the scientific Use Cases) and technically with European e-Infrastructures.

Subtask 4.1.1: Management and Coordination

Highlights

- Compute platform organised and Task leadership established
- Coordination tasks, meetings and presentations

Following a kick-off meeting in September 2015, in PY1 WP4 has seen the nomination of task leads for the ELIXIR Compute Platform, iteratively reviewed the scientific use cases (WP6 to 9) to generate requirements, and formed a community of experts from organisations participating in WP4 to start delivering on these requirements. An initial Technical Services Roadmap developed before the start of the project[12] formed the basis of these discussions with the scientific use cases. A prioritisation meeting in January 2016 resulted in an initial focus on non-human data and the core technical building blocks around AAI, Cloud and Data Movement. The 1st annual meeting of ELIXIR Compute Platform experts was held in Helsinki in June 2016. The 2nd annual meeting is planned to be held in June 2017. The task leads from across WP4 meet every 2 to 3 weeks to review individual progress, identify common issues, and plan future integrating work activity.

The current challenge for the ELIXIR Compute Platform is in receiving and interpreting technical requirements from the scientific use cases. In order to facilitate the interaction between science and technology participants, WP4 has created the concept of Technical Use Cases or TUCs. TUCs translate the engineering challenges faced by the driving scientific use cases in terms that are understandable by both parties. The first version of the ELIXIR Technical Services Roadmap, designed to drive forward life-science data analysis activities in Europe, focuses on identified TUCs and describes solutions created by the WP4 experts in the ELIXIR Nodes in collaboration with the European e-Infrastructure.

In the following sections an assessment is made of the implementation maturity of individual TUCs. The following definitions are in use in WP4, and build upon the definitions defined by Task 3.1 in WP3, to be defined as:

- **Proof of Concept (PoC) Services**: These are best effort services usually delivered to a few early-adopting users that are being actively co-developed to explore the functional capability in a particular area. PoC Services may

[12] https://drive.google.com/open?id=0ByWHBDVpEowoTm92N0VEUnV2UXc

Horizon 2020 grant n. 676559

have lower reliability compared to mature services and go through more changes in their presentation and APIs. There is no commitment to maintain the service and it could be withdrawn at very short notice.

- **Emerging Services (Prototype):** Emerging Services may have lower reliability compared to mature services and go through more changes in their presentation and APIs. The 'Emerging' status should not exceed 2 years. If an Emerging Service does not become Mature, it is good practice to display an "end of service" date prominently for at least 6 months before the service is withdrawn.
- **Mature Services (Production):** A Mature Service is active and reliable. If feasible, major changes to its APIs and/or user interface that may break existing functionality and/or are not fully backwards compatible are notified at least 6 months beforehand. A Mature Service relies only on other Mature or Legacy Services. In exceptional cases, a Mature Service might rely on an Emerging Service that is close to becoming mature. It is good practice that withdrawal of the Service is notified at least 1 year beforehand; during this period the Service has Legacy Status.
- **Legacy Services:** A previously Mature Service scheduled for retirement. A Service must spend at least 1 year in the Legacy state before final withdrawal. Reliability should be at the same level a Mature Service, but only the most serious issues are likely to be addressed and no new features will be added.

ELIXIR Compute Platform services map into the ELIXIR Service lifecycle as detailed in Table 9.

Table 8 Mapping of Compute Platform services to the ELIXIR Service lifecycle

ELIXIR Service Maturity	ELIXIR Compute Platform Service Maturity	Change management by the Compute Platform
	Proof of concept (Made available to early adopters)	decide on changes with small number of identified early adopters. Best effort service level.
Emerging	Emerging (Prototype in service)	decide on changes but provides notice of upcoming changes. Best effort service level.
Mature	Mature (Production in service)	coordinate changes with the relying ELIXIR Services. High availability is possible.
Legacy	Legacy (Production but being decommissioned)	coordinate termination date with relying ELIXIR Services.

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

Highlights

- Use of the Life Science Identity to access EGI's Operational and User Tools
- Collaboration with EUDAT 2020 around the Data Set Distribution Tool
- Development of a more streamlined deployment model for the EGI Federated Cloud based on feedback from ELIXIR Competency Centre.

European e-Infrastructures (e.g. GÉANT, EGI, EUDAT, PRACE) 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.

Cooperation with EGI over the first project year has been excellent at both a technical and managerial level. Some of this work has taken place through the ELIXIR Competency Centre within the EGI-Engage project[13],[14]. A significant amount of planned effort by EGI in allowing their operational tools to use federated identities (rather than X.509 certificates) has benefited ELIXIR. The EGI AAI gateway now links GoCDB, AppDB and in the future other EGI tools, to the ELIXIR AAI.

The EUDAT2020 project has been supporting the development of the Data Set Distribution Service to support use cases coming primarily from ELIXIR. However, similar use cases have also been identified in the environmental domains. The Data Set Distribution Service understand the structures of data sets and their releases, allowing a data set release to be defined as a collection of file sets that need to be distributed to distributed sites from a central authoritative source. The service uses this metadata to organise the movement of individual files to individual sites as required using established file transfer mechanisms. More details can be found in the description of task 4.3.3. Storage and Transfer.

While contact points have been established with GÉANT, GÉANT's Internet backbone and eduGAIN service are consumed indirectly by the ELIXIR Compute Platform. GÉANT's networking services are normally utilised through each ELIXIR Node's local networking provider, and it is through this networking provider that sites would escalate performance issues to GÉANT rather than contacting GÉANT directly.

[13] M6.3 milestone document: Life science requirements analysis and driver use case(s) with implementation roadmap (end of Feb 2016): https://documents.egi.eu/document/2675
[14] D6.10 deliverable document: Infrastructure tests and best usage practices for life science service providers (June 2016): https://documents.egi.eu/document/2841

Table 9 Technical Use Cases (TUCs) supporting the use European e-infrastructure services for ELIXIR

TUC ID	Status	Comment
9 - Infrastructure Service Directory	PoC	EGI provides an Infrastructure Service Directory (GoCDB) as part of its operational tools suite. This tool has been integrated to the ELIXIR AAI through EGI's new AAI gateway. ELIXIR sites, both within the ELIXIR Operations Centre (see Figure 6) and those in established national structures can be identified by an 'ELIXIR' tag that allows ELIXIR sites to be identified and used in other EGI Operational Tools (e.g. accounting and monitoring). Site managers and ELIXIR Operational staff are able to use GoCDB to register sites and update site details (Figure 7) and identify the services being run by a site. Currently, monitoring has only been enabled for services relating to EGI's Federated Cloud; however, discussions have started as to how the GA4GH beacons being deployed within ELIXIR could be integrated into EGI.
21 - Operational Integration	PoC	EGI's operational tools include a monitoring infrastructure. The cloud services deployed by CESNET and EMBL-EBI are tested at regular intervals using a Nagios-based system. Figure 8 shows the availability and reliability of the CESNET site within ELIXIR as part of this monitoring activity.
22 - Resource Accounting	PoC	For the cloud services integrated within ELIXIR using EGI's Federated Cloud model, consumption relating to cloud use can be recorded. More work is needed to establish the ELIXIR-related usage at these sites.

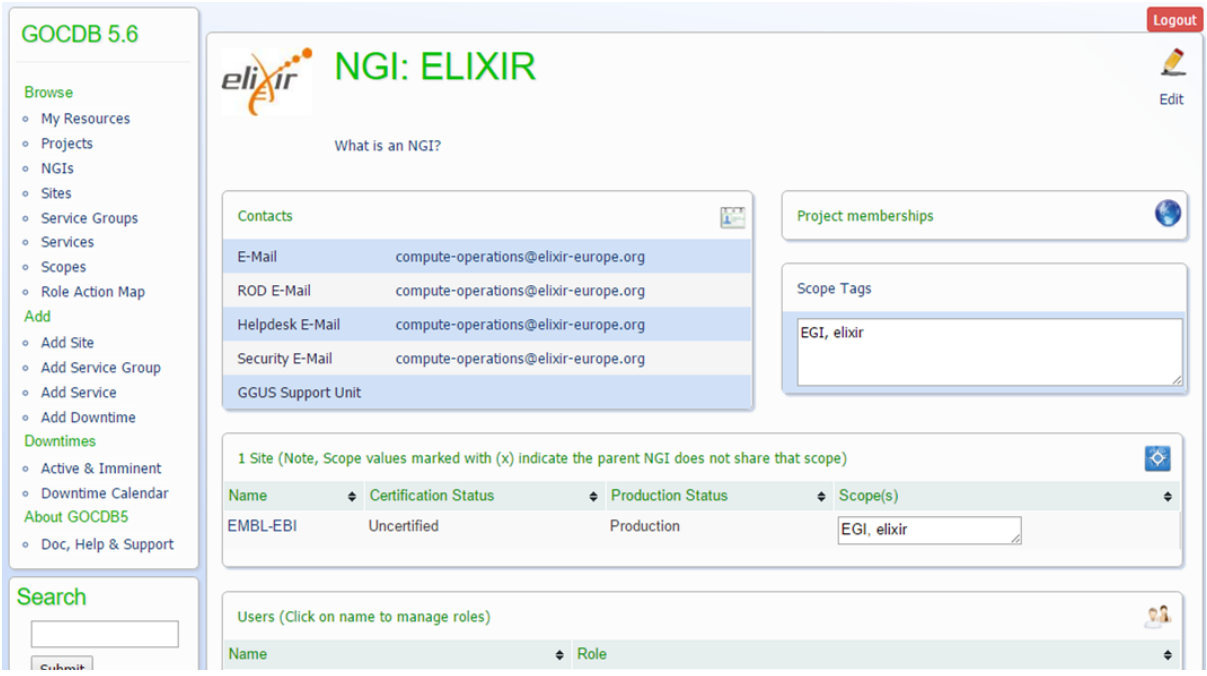

Figure 6 ELIXIR NGI within the EGI GoCDB tool that collects operational contact information and affiliates sites to an operational centre

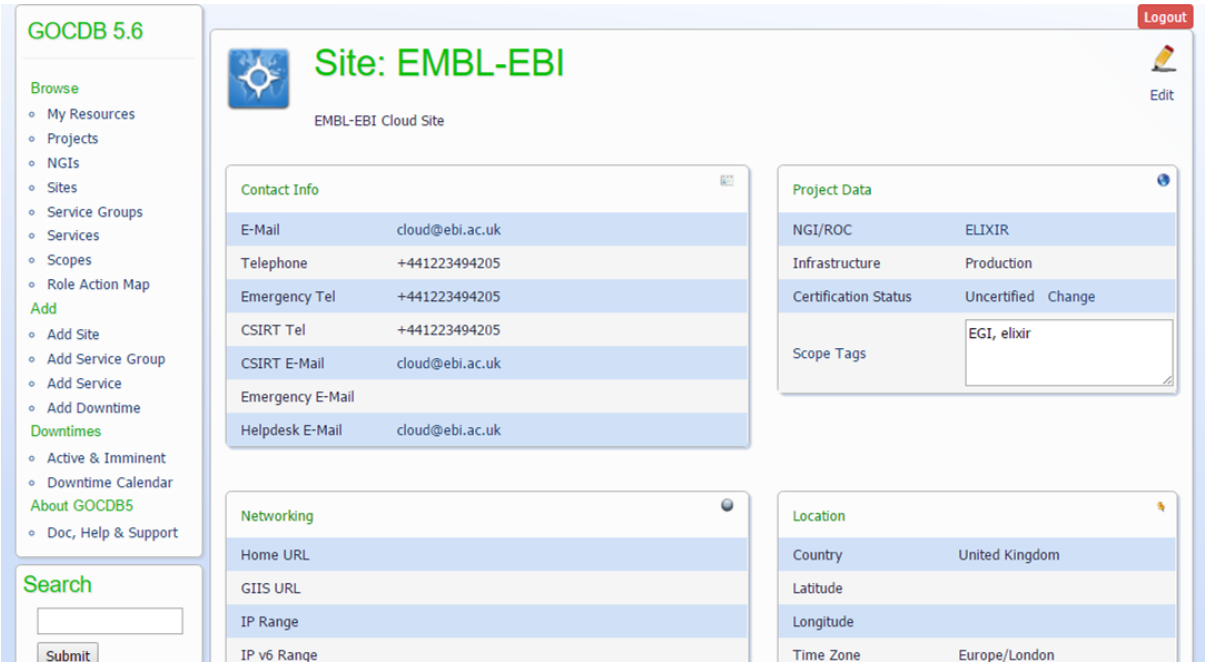

Figure 7 EMBL-EBI site within the GoCDB tool that is part of the ELIXIR Operations centre

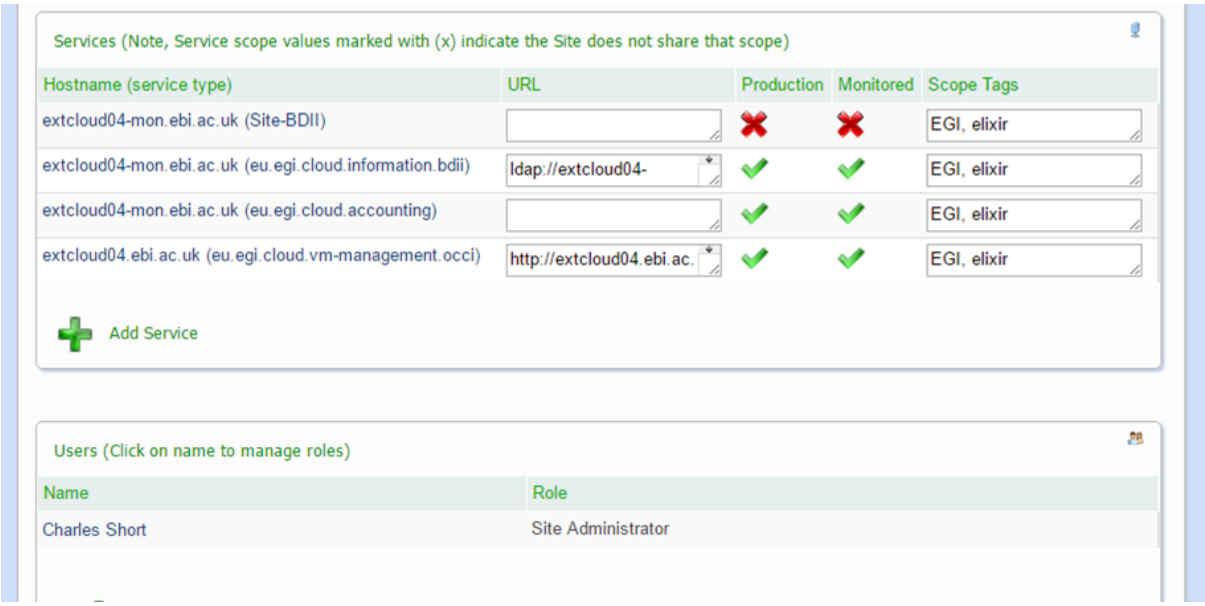

Figure 8 Services run at EMBL-EBI that are available for collaborative use within ELIXIR and which will be monitored by the EGI monitoring system

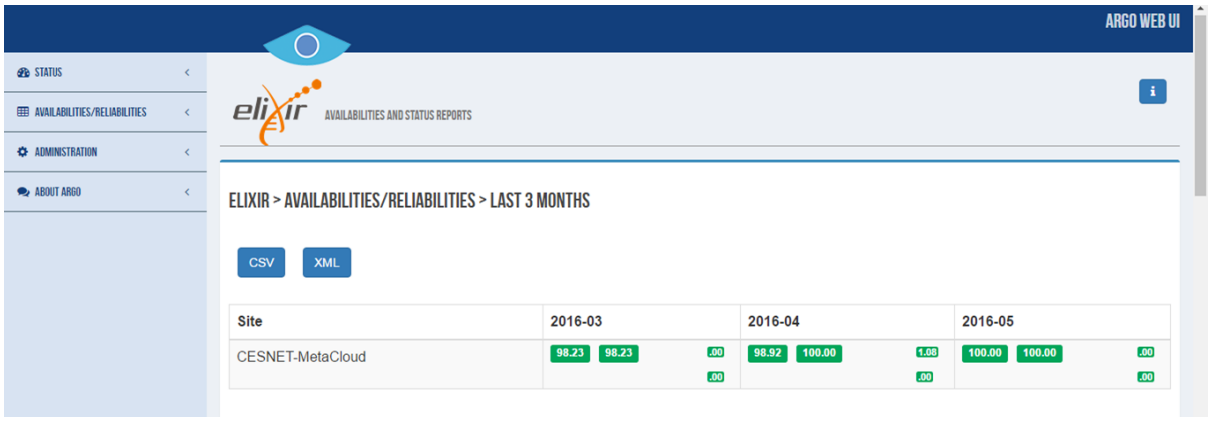

Figure 9 Monitoring of CESNET site services yielding their monthly availability (first number on the left) and reliability statistics (second number on the left)

Subtask 4.1.3 ELIXIR technical community building and knowledge exchange

Highlights

- ELIXIR Compute expert colleague network established
- EGI and GÉANT e-Infrastructures established dedicated contact points for ELIXIR

Since the start of ELIXIR-EXCELERATE, two major workshops were held in London (September 2015, January 2016) to organise the community of technical experts in ELIXIR Nodes into a Compute Platform. The WP4 community participated as an organised network for the first time in the ELIXIR all-hands meeting in Barcelona in March 2016.

ELIXIR Compute Platform core tasks in WP4 are the Authentication & Authorisation Infrastructure (AAI), Cloud & Compute and Storage & Transfer and Support. Tasks are co-chaired by two or three ELIXIR node experts and can connect with experts in the Nodes participating in WP4 to deliver on a specific TUC. Individual tasks hold focused meetings, write documentation, and follow ongoing actions on a regular (weekly) basis. Documentation is organised in a project-specific Google drive folder. The main documents are also available on the ELIXIR intranet.

The activities of the ELIXIR AAI task have already had an impact within European e-Infrastructures such as EGI and EUDAT. User identification and service access authorisation processes from these e-Infrastructures are compatible with the ELIXIR AAI, promising easier access in the future to these e-Infrastructures for European life scientists. EGI is already accepting ELIXIR AAI credentials across some of its services and EUDAT plans to introduce this capability in the future.

Task 4.2 User-Facing Support

The main consumers of this task, which establishes the user-facing support mechanisms within the ELIXIR Compute Platform, are the ELIXIR-EXCELERATE use cases (WP6 to 9) and other ELIXIR activities (e.g. other 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

- First iteration of scientific use case requirements was completed
- TUCs concept created to overcome challenges for communication between experts in Compute and Scientific Use Cases

This task collectively gathers requirements for the Compute Platform from the scientific use cases and prioritises their delivery through the technical services available within the ELIXIR Nodes. Invitations and questionnaires were sent to the use case WPs with the goal of reaching a set of common and unique requirements by end of the year 2015. It turned out to be challenging to establish technical discussions with the scientific use case experts as they were still trying to establish their own technical architecture. Thus a common and unique requirements analysis from the scientific use cases list is not yet completed.

Therefore, WP4 established the concept of Technical Use Cases (TUCs). The aim of these TUCs is to help translate the engineering challenges faced by the scientific use cases in terms that are understandable by both parties, as well as provide guidance for the implementation of technical solutions. In the autumn of 2015, WP4 decided that that the Compute Platform tasks can and should communicate their task-specific requirements directly using TUCs as translators without necessarily involving Task 4.2.1: the working model was changed from the initial model accordingly.

For PY2, the WP4 Tasks and the Scientific Use Case contact points will keep in touch, regularly reviewing their requirements as the collaboration matures. WP4 will prioritise the work based on the readiness and urgency expressed by Scientific Use Cases towards the Compute platform services. WP4's Technical Services Roadmap document will be updated iteratively as new requirements and solutions emerge. The planned work of this Task is invested in collaboration with the other Compute platform tasks due to changes in the working methodology, focusing on finding commonalities from requirements expressed by the Scientific Use Cases.

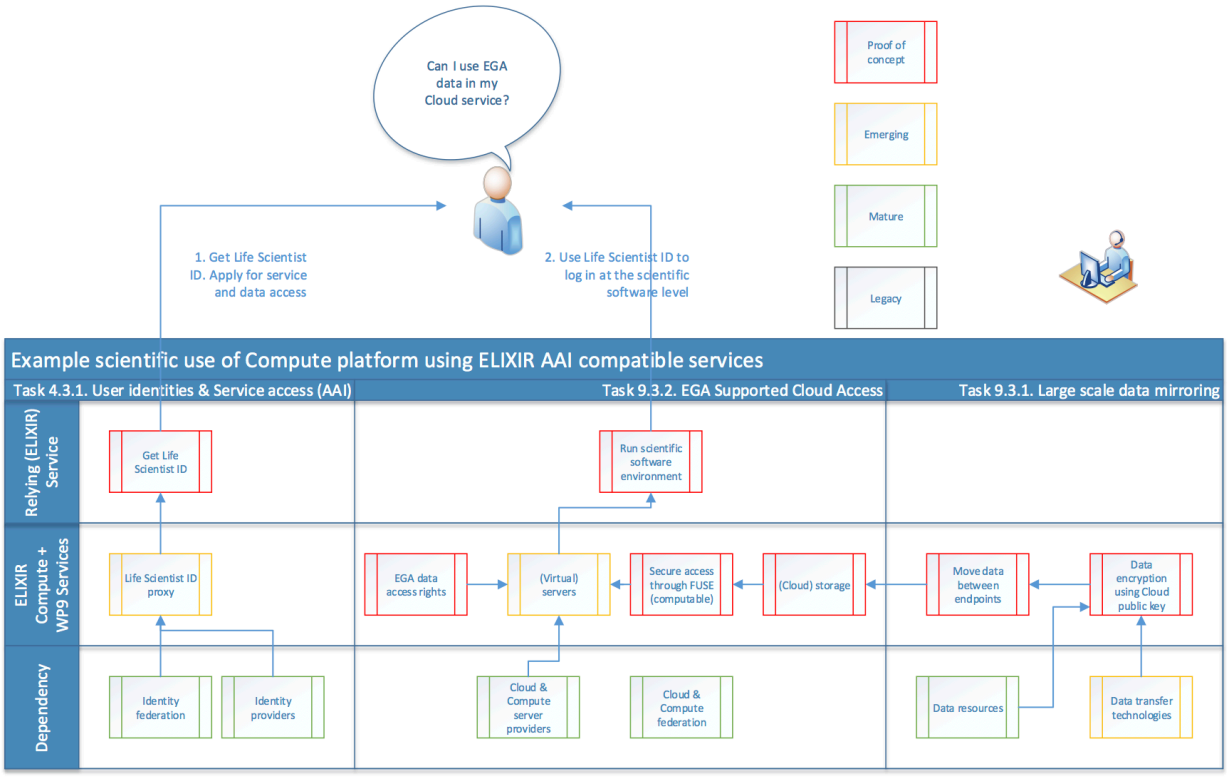

Figure 10 WP9 use case on authorizing data access from the EGA to a cloud service where datasets have been mirrored and stored in a secure manner

Subtask 4.2.2 User support and integration

Highlight

- Email contact points established, task ensures requests are tracked

In PY1 the focus was on understanding the existing helpdesk infrastructures within the ELIXIR partners so that current best practice could be recognised.

Over the last five years a support desk has been established in The Netherlands as part of a project that provides software infrastructure for health studies. This support desk has dedicated support personnel that handles first line support (issues with accounts and access, and reports about system disturbance). For issues that cannot be handled by the first line support team, a system of second line and third line support is in place.

It is planned that the experience gained in The Netherlands will also be applied in the ELIXIR-EXCELERATE project and, if this is successful, an extension to other ELIXIR Services is possible. During the second year of the project, information relating to the current Infrastructure Services will be collected in sufficient detail to support a first line help desk.

The help desk system in the NL, which is also used for capturing the Standard Operating Procedures for deployment and support, will be evaluated for use by ELIXIR and the feasibility of coupling this help desk to existing ticketing systems of ELIXIR Service providers will be investigated.

While this development is ongoing a lightweight help desk is available to users of the ELIXIR Compute Platform through a single mail address (contact-compute@ELIXIR-europe.org).

Task 4.3 Technical Infrastructure Integration

Task 4.3 integrates existing software and e-Infrastructure solutions to support the implementation of the ELIXIR Compute Platform. 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 of 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.

Underpinning all Infrastructure Services is the ELIXIR AAI, which provides a consistent ELIXIR identifier across the community that can be adopted by any of the Infrastructure Services (e.g. cloud, compute, storage) either directly or by using one of the credential translation mechanisms.

Subtask 4.3.1 ELIXIR AAI - Authentication, authorization (access)

Highlights

- Establishing the ELIXIR ID and related basic AAI services
- ELIXIR Heads of Nodes agreed on the ELIXIR AAI strategy
- Integrating the ELIXIR AAI with prototype credential translation services being developed by the AARC project
- Established contacts to the AAI experts in other biological and medical sciences research infrastructures in the context of the CORBEL and AARC projects

The ELIXIR AAI has been driven primarily from the ELIXIR-FI and ELIXIR-CZ Nodes with strong support and feedback from across WP4. Successful collaborations have been established with the AARC project in terms of providing requirements and benefiting from some of their service prototyping.

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.

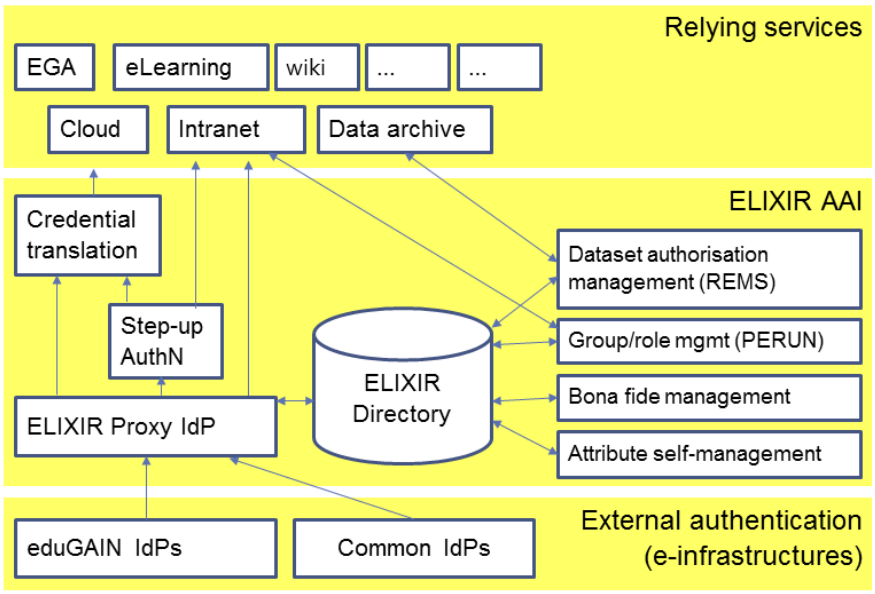

Figure 11 ELIXIR AAI design

Table 10 Technical Use Cases (TUCs) supporting authentication and authorization access

TUC ID	Status	Comment
1 - Federated ID	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	Emerging	Supported internet identity providers: Google, ORCID.
3 - ELIXIR Identity	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 within.
10 - Credential Translation	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	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	Intention to focus on this TUC together with WP9 in Project Year 2.

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 recognised services that can potentially be integrated to ELIXIR Cloud & Compute

During PY1 there has been a focus on the Cloud aspects of this task in order to meet the immediate needs of the scientific use cases. Individual ELIXIR Nodes have been establishing their cloud services. 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, as well as with e-Infrastructure activities such as EGI, Helix Nebula, and the European Open Science Cloud. The federation of ELIXIR cloud services has started with the creation of an ELIXIR Operations Centre within the EGI Operational Tools and the registration of the EMBL-EBI site alongside the established CESNET site. More details are provided in Section 4.1.2

Table 11 Technical Use Cases (TUCs) supporting Cloud and Compute integration

TUC ID	Status	Comment
4 - Cloud IaaS Services	Complete	Available information on the cloud services within ELIXIR has been collected from the ELIXIR Compute Platform partners[15] and will be made available on the ELIXIR website.
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	PoC	The EGI AppDB has been integrated with the ELIXIR AAI through the EGI AAI gateway and provides a repository for virtual appliances potentially available for the ELIXIR community[16]. Sites that are part of the EGI Federated Cloud are able to benefit from the vmcaster/vmcatcher software to distribute virtual appliances to sites before they are required, reducing startup times.
13 - Container Library	PoC	The BioShadock[17] service is provided by the French node but is not currently integrated with the ELIXIR AAI.

15 https://docs.google.com/document/d/1Dsb_O_EPYrof8WmVtNhBWc5aCnOqjkimdAIbSxYrB0k/edit#

16 https://appdb.egi.eu/store/vo/vo.elixir-europe.org

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

Horizon 2020 grant n. 676559

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	PoC	EGI's Federated Cloud technology has been tested by a subset of the ELIXIR nodes (CESNET and EMBL-EBI). Feedback from EMBL-EBI to EGI on the manual deployment process has led to a virtual appliances being developed by EGI that would greatly simplify the adoption of the EGI model by ELIXIR sites. The integration of the ELIXIR AAI with the X.509 credential translation services enables X.509 proxy certificates to be generated to access the EGI Federated Cloud services. EGI Federated Cloud sites are also integrated into the GoCDB (TUC21).
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

- Deploying GridFTP servers at ELIXIR sites and their integration and use through the ELIXIR AAI credential translation service
- Establishing a 'heartbeat' service to explore the performance of file transfers between different sites within the ELIXIR Compute Platform

The first year for subtask 4.3.3 Storage and Transfer has mainly been devoted to surveying and testing different technical solutions.

Additional information on requirements and plans for the scientific use cases were collected in a questionnaire and discussed in an online meeting in November 2015. As several of the scientific use case work packages were still in an early planning phase, a follow up to this will take place during second half of 2016.

To establish ELIXIR's distributed file transfer capability and enable network tuning, 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.

Contact has been established with other research infrastructures in order to collect information on existing technologies and services which could be integrated/used by the ELIXIR Compute Platform. This work, which is ongoing, includes collaborations and discussions with e.g. EGI, EUDAT and CERN.

For PY2, a pilot file transfer service will be launched that is integrated with the ELIXIR AAI system. Work will continue in collaboration with the EUDAT 2020 project around the Data Set Replication service (TUC15) and, depending on feedback from scientific use cases, Network File Storage (TUC7). In addition, coordination with WP5 (Interoperability) and WP9 (Human access-controlled data) will be strengthened, in particular with respect to metadata format and content, where the task at hand is mainly concerned with a subset of the metadata (data set definitions, PID:s, etc.).

Table 12 Technical Use Cases (TUCs) supporting storage and data transfer

TUC ID	Status	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	PoC	A test mesh of heartbeat file transfers has been established for testing the performance and robustness of the networking links and the file transfer services using them. These file transfer service make use of X.509 proxy certificates (TUC10) to authenticate transfers that are issued using the ELIXIR AAI. Documentation and deployment scripts for setting up a transfer endpoint have been produced and published on the ELIXIR intranet [LINK].
15 - Data Set Replication	PoC	Initial work on defining the metadata for data sets has taken place and an early prototype has been developed with support from EUDAT2020 and used to drive the heartbeat transfers. This work will continue in PY2 coordinating with WP5 Interoperability.
19 - PID and Metadata Registry	Later	This TUC is dependent on the data set metadata and will be coordinated with WP5.

The heartbeat file transfer mesh is composed of 6 active nodes (BILS - Sweden, CRG - Spain, FUNET - Finland, CSC - Finland, CESNET - Czech Republic & EMBL-EBI), with another two close to integration (SURFsara, Netherlands & GenOuest, France) and a further three in the pipeline (NeLS - Norway, Indiana University - USA & EGA - EMBL-EBI). It regularly transfers a file between all the combinations of nodes and records the transfer times and any failures. Since starting in May 2016, over 500,000 transfers have taken place within the first two months with a 91% success rate.

The data captured by the heartbeat mesh helps to understand the networking topology between the various ELIXIR Nodes. The actual numbers in the following graphs are not significant as they were captured during a prototyping phase but indicate the detail and the insight that can be collected through this service.

For example, Figure 12 shows the time taken by transfers, categorized by success/failure for each mesh edge, where low red dots indicate a fast successful transfer.

Figure 12 Time taken by transfers, categorized by success/failure for each mesh edge, where low red dots indicate a fast successful transfer

Another example for the successful transfers, is the more detailed timeline of durations as shown in Figure 13.

Horizon 2020 grant n. 676559

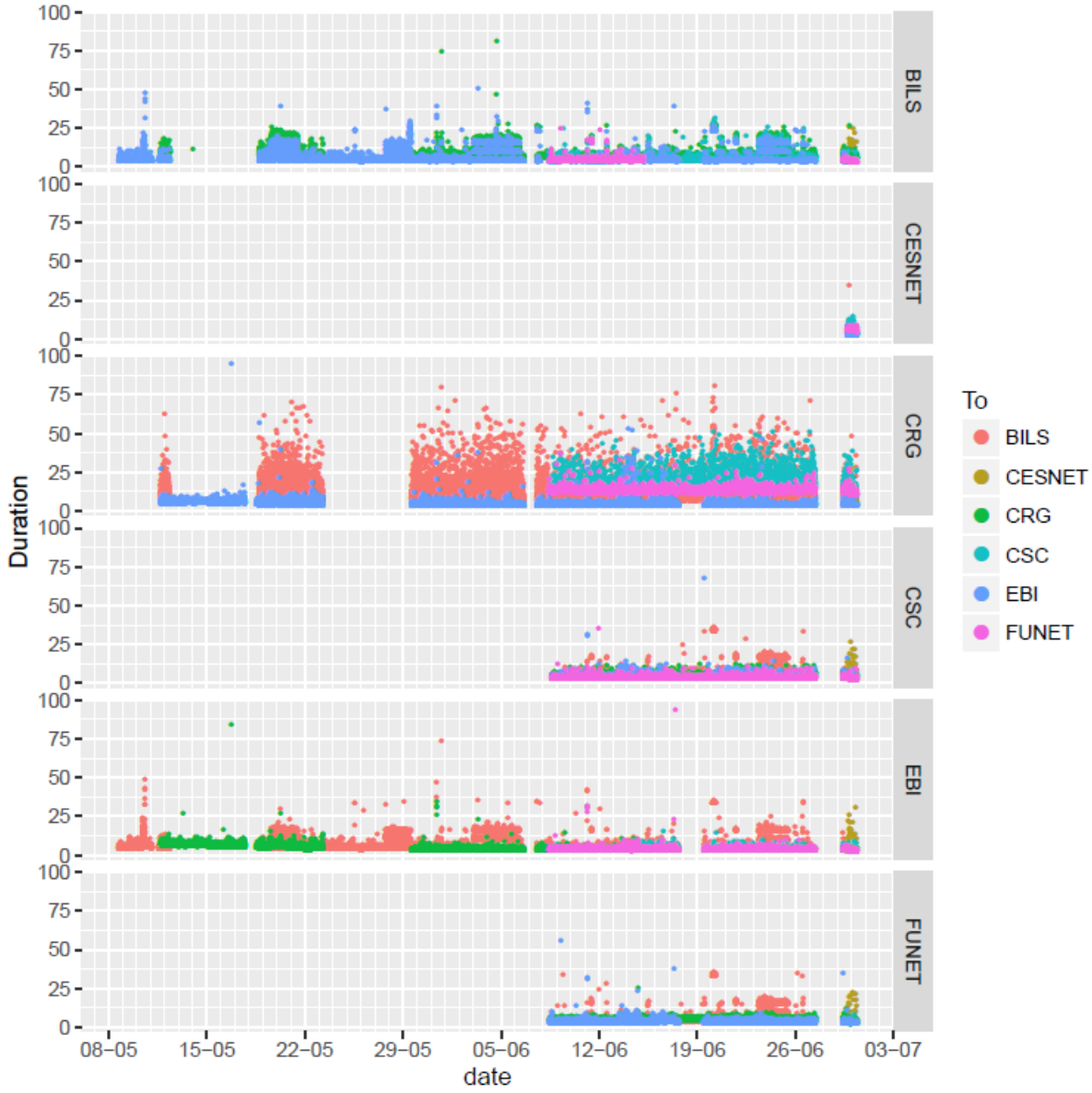

Figure 13 Detailed timeline of transfer durations

While the GridFTP transfers and the heartbeat tests provide the basis for understanding the backbone capability and capacity of data transfers within the ELIXIR Nodes, services based on top of this backbone still need to be defined. This will be taking place during PY2 and will explore:

- Customising data transfers to meet the needs of data centres and compute infrastructures to allow data handling according to specific data policies
- Using third-party data transfers services (e.g. webFTS/FTS & Globus Online) to centrally plan and monitor data transfers
- Triggering data transfers from within workflows using third-party data transfers.

Horizon 2020 grant n. 676559

Subtask 4.3.4 Infrastructure Service Registry

The Infrastructure Service Registry 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 Service 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 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.

In PY1 effort has been mainly focused in other areas of the Compute Platform as there are a number of mature technologies that could be adopted and adapted for this task. This will be the focus in PY2.

Table 13 TUCs supporting the Infrastructure Service Registry

TUC ID	Status	Comment
16 - Infrastructure Service Registry	Later	The priority during PY1 has been to establish the Infrastructure Services (e.g. Cloud, File Transfer) that would be recorded in the Infrastructure Service Registry for discovery by applications (programmatically) and by users (web interface). There are a number of technologies in use by the e-Infrastructure community (e.g. BDII or the ATLAS Grid Information System) or from within ELIXIR (e.g. ELIXIR Service Registry).

Horizon 2020 grant n. 676559

Appendix D: Analysis of Compute Platform Recommendations

During the development of the ELIXIR Technical Services Roadmap a number of recommendations were made. These recommendations are reviewed in this section to identify the progress made and to examine whether they are still relevant. Each recommendation is assigned to one of the following categories:

- Open: The recommendation exists but no practical work has been taken around it.
- Accepted: The recommendation has generally been accepted by the ELIXIR community as an approach to explore, possibly alongside others.
    - Ongoing: The work relating to this recommendation is ongoing.
    - Completed: The work relating to this this recommendation has been completed.
- Rejected: This recommendation is no longer relevant and is no longer being pursued.

Table 14 ELIXIR Compute Platform Policy Recommendations

ID	Recommendation	State & Analysis
R1	All organisations with researchers active within ELIXIR should attempt to deploy Institutional Identity Providers wherever possible and register it to eduGAIN.	Accepted & Completed. Individual high-quality institutional Identity Providers remains the ideal model, however the ELIXIR AAI model is not dependent on it.
R2	A security profile based around the ELIXIR AAI Model will be specified and maintained by the AAI Task Force (using SAML, OAuth2 and other specifications) and this profile must be supported directly by all ELIXIR Infrastructure Services (i.e. Storage, HTC/HPC Cluster, Cloud IaaS, …) requiring authenticated and authorised access, or they must provide/contribute to a credential translation mechanism that would allow indirect access.	Accepted & Completed. The ELIXIR AAI model has been implemented based around standards and is available for services to adopt and build upon.
R3	All ELIXIR Infrastructure Services must publish a unique resolvable identifier (e.g. a PID for the service) which users of that service can use to acknowledge the resources that they have used (e.g. in publications).	Open.
R4	As a condition of use, all use of ELIXIR Infrastructure Services known to a user should be acknowledged by researchers in their publications by referencing the services unique identifier together	Open

Horizon 2020 grant n. 676559

	with a verbose normal reference: i.e. name of the service, institution, maintainer.	

Table 15 ELIXIR Compute Platform Technical Recommendations

ID	Recommendation	State & Analysis
R5	An ELIXIR Identity (expressed as a group membership) can be associated with an external identity (e.g. Organisational Login, LinkedIn, Google, etc) and membership of the group will require a specified accreditation process (e.g. confirmation by two existing ELIXIR members or by a designated Node contact).	Accepted & Completed. The ELIXIR AAI supports multiple identity providers.
R6	An ORCID is expected to be one of the attributes associated with an ELIXIR ID that can be provided to relying parties. This can be self assigned by the user using the ORCID API (effectively an IdP) to demonstrate that they 'own' a particular ORCID.	Accepted & Ongoing. The implementation work is underway.
R8	ELIXIR should negotiate with the Globus for the provision of Globus Transfer and Globus Share capabilities built around the ELIXIR Identity model. If this negotiation is not successful then the equivalent capability can be delivered through other independent services that would have to be integrated together. If this negotiation is successful, and adoption of Globus Transfer and other services within ELIXIR increases, then consideration will be given for a longer-term relationship with the Globus team.	Accepted & Ongoing: The broader recommendation for file transfer is accepted. The ELIXIR AAI is currently being exposed through OIDC to provide an experimental integration with Globus.
R9	The PERUN software be used to establish the endorsed 'ELIXIR Researcher' attribute and provide Group Management functions within ELIXIR.	Accepted & Completed. PERUN is central part of the ELIXIR AAI.
R10	Infrastructure Service providers should deploy GridFTP (using the Globus Connect Server package) and make their hosted data resources available within the Globus Transfer Service for at least read-only access and read-write access for authorised users.	Accepted & Ongoing: GridFTP servers are being deployed and their effectiveness being assessed.
R13	Infrastructure Service Providers may make a local decision as to which of the underlying Cloud IaaS (e.g. OpenStack) or HTC/HPC Cluster technologies (e.g. LSF, Slurm) they deploy provided they are supported by an EGI Technology Stack (e.g. gLite, ARC, UNICORE) that is compatible with ELIXIR's AAI model (see the result of R12 - EGI Technology Stacks).	Accepted & Ongoing. So far only two cloud sites (based on OpenStack and OpenNebula) are being integrated using the EGI model.

R15	The BDII service (based on OpenLDAP) used within EGI should be adopted as the ELIXIR Infrastructure Service Registry and deployed centrally given its proven ability to scale. The use of the ELIXIR Service Registry to register and discover Infrastructure Services is seen as being outside of its current development scope.	Open. Work has not yet started on the ELIXIR Infrastructure Service Registry.
R16	All ELIXIR Infrastructure Services (File Transfer, Network File Storage, Cloud IaaS, Cloud Storage and HTC/HPC Cluster) must be registered in the ELIXIR Infrastructure Service Registry so that they can be discovered by users.	Open. Work has not yet started on the ELIXIR Infrastructure Service Registry.
R18	ELIXIR should use the EGI Federated Cloud model as the approach to integrate its Cloud IaaS alongside others.	Accepted & Completed: One new cloud site and one existing cloud site have been working with EGI's Federated Cloud model.
R19	ELIXIR should use the EGI High Throughput Computing model as the approach to integrate its HTC/HPC Clusters alongside others.	Open. No ELIXIR sites have plans to federate their clusters.
R20	ELIXIR should use the EGI Applications Database for ELIXIR to distribute Virtual Machine Images to ELIXIR sites within the EGI Federated Cloud. The EGI Applications Database must provide an API that allows information relating to its public content to be integrated into other systems (e.g. ELIXIR Tools Registry). EGI should consider how data from other sources within ELIXIR relating to Virtual Machine Images (e.g. Software Tool Registry) could be integrated into the EGI Applications Database.	Accepted & Completed.
R22	Infrastructure Services and Infrastructure Service Providers should adopt and integrate with the REMS software to provide data owners (or their representatives, such as Data Access Committees) an opportunity to manage access rights to restricted data sets (e.g. human related) and analysis resources.	Open. No demand for adoption yet by Infrastructure Service Providers.
R23	The ELIXIR Compute Platform should adopt the EGI GoCDB to hold information on the ELIXIR Infrastructure Services that are expected to be operating within ELIXIR by each Node.	Accepted & Completed: The GoCDB has been integrated with the ELIXIR AAI.
R24	The ELIXIR Compute Platform should adopt the EGI APEL model to collect individual Infrastructure Service usage information by any individual (and thereby their affiliated project/group) that is authenticated using the ELIXIR ID.	Accepted & Ongoing: The ELIXIR sites within the EGI Federated Cloud are using this

		model.
R25	The ELIXIR Compute Platform should adopt the EGI ARGO service to provide service level monitoring and reliability information for ELIXIR Infrastructure Services.	Accepted & Ongoing: The ELIXIR sites within the EGI Federated Cloud are using this model.
R27	The ELIXIR Metrics Portal should use the EGI APEL service through a programmatic API as the source of its ELIXIR related usage based reports.	Open. Waiting for APEL to be fully utilised by ELIXIR.

Table 16 ELIXIR Compute Platform Activity Recommendations

ID	Recommendation	State & Analysis
R7	A survey of ELIXIR Infrastructure Service providers should be undertaken to see if an additional attribute associated with an ELIXIR ID that could relate to a 'step-up' authentication mechanism would be of general use. If only relevant to human data then this could be embedded as part of the human data entitlement authorisation process. However, it could also be required to use specified administrative functions and could be enabled for all account access if an individual desired.	Accepted. The mechanism that will be used to provide 'step-up' authentication is still being discussed.
R11	A Data Transfer Task Force should be set up immediately as part of ELIXIR-EXCELERATE with membership from DANTE/NRENs to provide recommendations as to best practice deployment and configuration of GridFTP and related services (e.g. WAN VPNs and secure connections), and explore transfer speeds and bottlenecks between key ELIXIR sites using the aforementioned technologies.	Accepted & Completed. A data transfer task force has been set up as part of T4.3.3 to deploy GridFTP servers. Links have been established with NRENs as required.
R14	The ELIXIR Cloud TF should work with EGI and other service providers to better understand the available CDMI implementations (e.g. dCache, Emotive, OneData) and other implementations to better understand the Technical Use Case and the solutions that are available.	Open. Work has not started
R21	The Data Transfer TF should work with EUDAT to demonstrate the use of the B2SHARE service to register a source data set (and its releases) and then use the B2SAFE/B2STAGE services to push a tree-based replication of large data sets across Europe (by leveraging deployed GridFTP endpoints) so that replicas can be discovered using B2FIND. This should not require any further software to be installed at a participating service provider.	Accepted & Ongoing. Work is taking place within EUDAT2020 to develop a Data Set Distribution Service to meet this use case.

Table 17 ELIXIR Compute Platform Recommendations for External Service Providers

ID	Recommendation	State & Analysis
R12	The EGI should identify which of its technology stacks are directly or indirectly compatible with the ELIXIR AAI model and document how this is done.	Accepted & Completed. EGI has deployed an AAI Gateway which integrates with the ELIXIR AAI. EGI services are now being adapted to use the EGI AAI Gateway. No further action for ELIXIR.
R17	Other ELIXIR Infrastructure Service providers (e.g. GEANT Cloud Exchange or Helix Nebula) are expected to advertise their services in the ELIXIR Infrastructure Services Registry.	Open.
R26	EGI related management services (i.e. GoCDB and ARGO) must adopt the ELIXIR AAI model to govern access to their services.	Accepted & Completed: Work from ELIXIR & EGI has integrated the ELIXIR AAU into GoCDB and AppDB.

Horizon 2020 grant n. 676559

Appendix E: Status of Compute Platform Recommendations

Looking at the recommendations from the perspective of their status:

Open

R3: All ELIXIR Infrastructure Services must publish a unique resolvable identifier (e.g. a PID for the service) which users of that service can use to acknowledge the services that they have used (e.g. in publications).

R4: As a condition of use, all use of ELIXIR Infrastructure Services known to a user should be acknowledged by researchers in their publications by referencing the services unique identifier together with a verbose normal reference: i.e. name of the service, institution, maintainer.

R15: The BDII service (based on OpenLDAP) used within EGI should be adopted as the ELIXIR Infrastructure Service Registry and deployed centrally given its proven ability to scale. The use of the ELIXIR Service Registry to register and discover Infrastructure Services is seen as being outside of its current development scope.

R16: All ELIXIR Infrastructure Services (File Transfer, Network File Storage, Cloud IaaS, Cloud Storage and HTC/HPC Cluster) must be registered in the ELIXIR Infrastructure Service Registry so that they can be discovered by users.

R19: ELIXIR should use the EGI High Throughput Computing model as the approach to integrate its HTC/HPC Clusters alongside others.

R22: Infrastructure Services and Infrastructure Service Providers should adopt and integrate with the REMS software to provide data owners (or their representatives, such as Data Access Committees) an opportunity to manage access rights to restricted data sets (e.g. human related) and analysis resources.

R27: The ELIXIR Metrics Portal should use the EGI APEL service through a programmatic API as the source of its ELIXIR related usage based reports.

R14: The ELIXIR Cloud TF should work with EGI and other service providers to better understand the available CDMI implementations (e.g. dCache, Emotive, OneData) and other implementations to better understand the Technical Use Case and the solutions that are available

R17: Other ELIXIR Infrastructure Service providers (e.g. GEANT Cloud Exchange or Helix Nebula) are expected to advertise their services in the ELIXIR Infrastructure Services Registry.

Accepted

R7: A survey of ELIXIR Infrastructure Service providers should be undertaken to see if an additional attribute associated with an ELIXIR ID that could relate to a ‘step-up’ authentication mechanism would be of general use. If only relevant to human data then this could be embedded as part of the human data entitlement authorisation process. However, it could also be required to use specified administrative functions and could be enabled for all account access if an individual desired.

Accepted & Ongoing

R6: An ORCID is expected to be one of the attributes associated with an ELIXIR ID that can be provided to relying parties. This can be self assigned by the user using the ORCID API (effectively an IdP) to demonstrate that they 'own' a particular ORCID.

R10: Infrastructure Service providers should deploy GridFTP (using the Globus Connect Server package) and make their hosted data resources available within the Globus Transfer Service for at least read-only access and read-write access for authorised users.

R13: Infrastructure Service Providers may make a local decision as to which of the underlying Cloud IaaS (e.g. OpenStack) or HTC/HPC Cluster technologies (e.g. LSF, Slurm) they deploy provided they are supported by an EGI Technology Stack (e.g. gLite, ARC, UNICORE) that is compatible with ELIXIR's AAI model (see the result of R12 - EGI Technology Stacks).

R24: The ELIXIR Compute Platform should adopt the EGI APEL model to collect individual Infrastructure Service usage information by any individual (and thereby their affiliated project/group) that is authenticated using the ELIXIR ID.

R25: The ELIXIR Compute Platform should adopt the EGI ARGO service to provide service level monitoring and reliability information for ELIXIR Infrastructure Services.

R21: The Data Transfer TF should work with EUDAT to demonstrate the use of the B2SHARE service to register a source data set (and its releases) and then use the B2SAFE/B2STAGE services to push a tree-based replication of large data sets across Europe (by leveraging deployed GridFTP endpoints) so that replicas can be discovered using B2FIND. This should not require any further software to be installed at a participating service provider.