# EXCELERATE Deliverable D10.2

| | | |
|---|---|---|
| **Project Title:** | ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences | |
| **Project Acronym:** | ELIXIR-EXCELERATE | |
| **Grant agreement no.:** | 676559 | |
| | H2020-INFRADEV-2014-2015/H2020-INFRADEV-1-2015-1 | |
| **Deliverable title:** | Blueprint on how to set up an ELIXIR data Node | |
| **WP No.** | 10 | |
| **Lead Beneficiary:** | UU | |
| **WP Title** | ELIXIR Node Capacity Building and Communities of Practice | |
| **Contractual delivery date:** | 31 August 2019 | |
| **Actual delivery date:** | 29 August 2019 | |
| **WP leader:** | Bengt Persson<br>Jiri Vondrasek<br>Brane Leskosek | 45 - UU<br>34 - UOCHB<br>32 - UL |
| **Partner(s) contributing to this deliverable:** | UU (SE), CSC (FI), UiO (NO), CRG (ES) | |

**Authors and Contributors:**
Niclas Jareborg, UU; Bengt Persson, UU;

**Reviewers:**
Gary Saunders (ELIXIR Hub)

# Table of contents

# 1. Executive Summary

This report describes how to set up a local/federated node in EGA (European Genome-phenome Archive). Human biomedical research data is often sensitive in nature, and projects consequently have strong requirements on IT security, data access control, and data management. The rapid increase in human genome sequence data, and the heterogeneous legal landscape among different countries for this type of data, makes it evident that a federated EGA organisation is the optimal solution for the future.

Within ELIXIR-EXCELERATE, we have worked on producing portable code that can be deployed at sites that want to be able to archive sensitive human genetic data locally, while still being discoverable and accessible through the central EGA interfaces, for researchers that have legitimate reasons for accessing the sensitive data.

A local/federated EGA node will be instrumental for each ELIXIR node in storing and making available sensitive research data produced in the respective country, enabling re-use of human genetic data in a protected environment that only allows legally correct access in accordance with ethical permits of the respective scientific studies. Thus, the local EGA node will make genetic data FAIR (Findable, Accessible, Interoperable and Reusable).

The local/federated EGA forms a sustainable, secure and legally correct storage of sensitive human genetic data, which facilitates data publication and open science, which in turn enables data sharing of benefit for scientists, healthcare providers and industry. Furthermore, access to European data (and in the future global data) accelerates life sciences and paves the ground for new directions of research. This access to human data is of special importance for healthcare providers when treating patients with rare diseases having genetic variants that are only known in a few cases world-wide.

# 2. Impact

We expect the local/federated EGA to be instrumental for large cohorts of human samples from research and national healthcare initiatives, especially for multi-national ones. Furthermore, the local/federated EGA will be an important infrastructure for the EU member states declaration to sequence and share transnationally at least 1M human genomes by 2022[1].

# 3. Project objectives

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

| No. | Objective | Yes | No |
|---|---|---|---|
| 1 | Implement a programme of organisational capacity building in newly formed ELIXIR Nodes, including sharing of best practice between partners in accessing EU Structural Funds (ESIF) for operating infrastructure | | X |
| 2 | Construct and coordinate ELIXIR-wide 'communities of practice' that support and develop the professionals who deliver advanced data and bioinformatics support and services in ELIXIR Nodes | X | |

# 4. Delivery and schedule

The delivery is delayed: Yes • No☑

[1] http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=50964

# 5. Adjustments made

N/A

# 6. Background information

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

| | | | |
|---|---|---|---|
| Work package number | 10 | Start date or starting event: | month 1 |
| Work package title | **ELIXIR Node Capacity Building and Communities of Practice** | | |
| Lead | Bengt Persson (SE), Jiří Vondrášek (CZ) and Brane Leskosek (SI) | | |
| **Participant number and person months per participant**<br>1 – EMBL 6.00, 2 – UOXF 4.00, 5 – UTARTU 20.00, 7 – CNIO 1.00, 9 – CIPF 3.60, 13 – CSIC 2.00, 16 – FCG 2.00, 17 - INESC-ID 10.00, 20 – CSC 4.00, 21 – UiB 4.00, 23 – UiT 4.00, 26 – CNRS 5.00, 31 – LIU 24.00, 32 – UL 30.00, 34 – UOCHB 8.00, 35 – MU 26.60, 37 – VIB 10.00, 39 – BSRCAF 12.00, 40 – HUJ 8.00, 42 – FORTH 6.00 | | | |
| This WP will address the issue of how to get people in Nodes coming together in capacity building, as detailed in the tasks below. There will be accompanying training needs in this capacity building and those training needs will be addressed in WP11. The training needs are in advanced training of the staff handling data and performing genome annotation and assembly. Other training needs for Use Cases will be in general addressed in WP11, but not specific to every Node. For Node capacity building, advanced training will be needed also in management and know-how on operating Nodes, performed in close collaboration with Task 10.1.<br><br>A Community of practice is a group of people who share a craft or a profession, created to coordinate efforts to solve defined tasks and/or with the goal of gaining knowledge related to their field. ELIXIR is looking to establish such Communities of Practice of bioinformatics experts involved in advanced bioinformatics user support across the Nodes to effectively interact with bioinformatics infrastructure users at interfaces of different research fields. ELIXIR Communities of Practice would be the primary mechanism for ELIXIR to establish domain specific services, for example, forming a community of genome annotators across Nodes to meet the need from national researchers of ready access to genome annotation resources. Other examples could be to meet the needs of Rare Disease or Medical genomics research, agricultural or marine bioinformatics and chemical compounds for biology. | | | |

ELIXIR will start to build these Communities of Practice to enable coordination and knowledge exchange in selected areas in tasks 10.2 and 10.3. Task 10.2 is directed to create Good Practices in setting up data Nodes, of importance to create a sustainable and scalable data flow from laboratories to national Nodes and further to European or global databases. Task 10.3 is directed to coordinate and exchange expertise in the field of genome annotation and assembly and to create Good Practice in for this field. In the future, further communities of practice are envisioned, arising from needs identified by the Use Cases (WP6 to 9) and identified through community workshops and surveys (Task 10.4). The creation of a sustainable mechanism for establishment of communities of practice is also addressed in Task 10.4.

**Objectives**

WP10 is focused on strengthening the ELIXIR infrastructure by supporting coordination of Node activities and increasing the organisational capacities of ELIXIR Nodes. ELIXIR Nodes are at very different levels of maturity, ranging from national infrastructures that have existed for over a decade to newly formed consortia. Activities will focus on spreading the knowledge and bioinformatics best practice that exists within ELIXIR's larger and more established Nodes, with newer or smaller ELIXIR Nodes in less research-intensive areas of the EU. This will help to create a stairway to excellence for partners involved, and support the creation of a true European Research Area. One of the deliverables will be a set of "Good practices" for setting up and running an ELIXIR Node, which will be of substantial value for both current and future Nodes.

Its two Objectives are:

1. Implement a programme of organisational capacity building in newly formed ELIXIR Nodes, including sharing of best practice between partners in accessing EU Structural Funds (ESIF) for operating infrastructure.
2. Construct and coordinate ELIXIR-wide 'communities of practice' that support and develop the professionals who deliver advanced data and bioinformatics support and services in ELIXIR Nodes.

Work Package Leads: Jiří Vondrášek (CZ) and Bengt Persson (SE)

Description of work and role of partners

**Task 10.1: ELIXIR Node Capacity Building (46PM)**

This task will support the formation of an ELIXIR community. There are significant differences between existing ELIXIR Nodes in their capacity, level of expertise and maturity of services/tools/data. We will increase the joint competence and capacity for Nodes lacking a large national user community, large-scale projects and big data or having a limited record of offered tools and services. These Nodes will benefit from mutual collaboration and connection with well-established and more advanced Nodes they can utilize their know-how for a more rapid Node development. Altogether, this will help shape ELIXIR as an efficient pan-European infrastructure.

The major aim of this task is to provide management knowledge transfer among Nodes to create a set of well-balanced, well-functioning and compatible Nodes.

Support in coordinating national Nodes, including Skills and Knowledge exchange between ELIXIR Nodes. Nodes with different experiences will help to provide knowledge regarding good practice in different situations and providing direct support to implementation of national infrastructures (e.g. by national / regional workshops with external experts, support to national community building efforts). The heterogeneity of Nodes established will help providing multiple effective ways for coordination and to get funds from national providers and their commitments to the infrastructure. Knowledge exchange will be catalysed by workshops, staff exchange programme and visits. This activity is based on the ELIXIR community practice experience but it is more general and should cover some features brought by larger staff community.

Identify and apply technical solutions at/between Nodes. The reason for particular technical solution must be explicitly formulated and the solution must be applicable on more than 2 Nodes. The capacity building deliverables would be primarily workshops based on Technical Services and/or Training WP deliverables.

Partners: CH, CZ, EE, NO, PT, SE, SI, UK, ES, EL, IL, EMBL-EBI

**Task 10.2: Capacity Building in Data Nodes Network (34PM)**

One of the aims of ELIXIR is to establish a network of data Nodes (Nodes with large data collections and databases with established way of data deposition and curation) to enable scalable data storage and their transferability by means of standardised formats. In this task, we will focus on establishing guidelines and good practices to facilitate efficient data collection into core data resources (cf. WP3), primarily focusing on data needed for selected Use Cases (WP6 to 9). This is tightly linked with IT solution by means of storage, dedicated networks and connections (cf. WP4). A distributed network following the same standards will also simplify international sharing of datasets for which this is ethically permitted.

This task both includes creation of routes for data publishing in a uniform manner across ELIXIR with data Nodes in each country and includes data repositories for replication of reference data allowing for fast access. The setting up of a data Nodes network has been identified by the technical experts within ELIXIR as a prioritised area.

Task 10.2 also includes development of Good Practices in setting up data Nodes enabling secure storage of sensitive data, such as sequence data related to patients. The task is interfacing with WP4 regarding technical developments on AAI and data transfer. Furthermore, there are connections with WP4 on data interoperability and the Use Case in WP9 on sensitive data.

Partners: SE, FI, CZ, EMBL-EBI, SI, PT, ES, EE. In due time, all ELIXIR Nodes are expected to have an ELIXIR data Node.

**Task 10.3 – Capacity Building in Genome Assembly and Annotation (44PM)**

Specialised expert platforms for genome assembly and annotation are already available in several ELIXIR countries. They provide critical support to complex genome projects and deliver annotations that serve as the basis for scientific inquiry into the genomics of newly sequenced organisms. The specialised expertise at multiple ELIXIR Nodes would benefit from capacity building through competence-spreading advanced workshops and staff exchange.

The capacity-building efforts will benefit the Use Cases in WP6 on marine organisms and in WP8 on plant Use Cases. The genome annotation groups will contribute with domain-specific knowledge about different species, e.g. marine organisms (SE, NO), woody plants (PT) and crop plants (SI).

Furthermore, in order to facilitate access to genome annotation to the users, we propose a deployment of web services to enable genome projects in the scientific community to efficiently interact with the data. The development of such web services is intended together with the EnsEMBL team to create a pan-European collaboration on genomics resources to provide researchers with a unified analysis platform carried by multiple partners.

Partners: SE, NO, FR, PT, EBI, SI, BE, CZ, ES.

**Task 10.4 – Sustainability of capacity building (30PM)**

The main goal of Task 10.4 is periodical and long-term discovery of users with specific capacity needs at ELIXIR Nodes and/or research groups within Nodes. This knowledge of capacity needs/gaps will be gathered through surveys and face- to-face meetings. With capacity needs identified the Task 10.4 team will connect users with WP11 groups that have at their disposal training infrastructure, learning materials and knowledge needed to implement the capacity building. In order to ensure the sustainable flow of knowledge and stable capacity maintenance we need to provide long-term networking of capacity seekers and providers. They will be focused to the great extent to the Good Practices from Task 10.2 and 10.3 (and WP6 to 9). With well-formed ELIXIR Communities of practice, the Task 6.4 will be able to lead the reuse or even suggest the adaptation of WP11 courses and training materials for specific capacity building needs.

It is of great importance that capacity needs will be periodically (but in long-term perspective) tested through surveys, which will also contribute to the sustainability of training infrastructure and learning materials provided by WP11. Task 10.4 will monitor the implementation of capacity building in Tasks 10.1, 10.2 and 10.3 in order to extract good practices and compile good practice recommendations and guidelines which can be used in other capacity building contexts.

Partners: SE, SI, CZ, BE, EE, EL, IL, EMBL-EBI

**Task 10.5: Supporting ELIXIR Nodes in understanding Smart Specialisation Strategies and accessing EU Structural and Investment Funds (ESIF) (36.2PM)**

The potential for exploiting funding synergies between EU Research programmes and ESIF are well known62. Those ELIXIR Nodes eligible for ESIF are therefore presented with a real opportunity for local funding of their Node, particularly in light of the proposed focus on ESIF and ESFRI that many Member States are making within their national plans to the Junker Investment Plan. However, understanding the local priorities for funding, rules, and application procedures presents is complex and time consuming and securing ESIF for operational costs of life science infrastructures is a real challenge. For ELIXIR Nodes to access ESIF in any meaningful way, support needs to be targeted at the local level, allowing scientists to build up an understanding of their local Smart

Specialisation Strategy, which dictates the funding opportunities for that region, and then develop a strong business case that can be used for subsequent funding applications.
Partners: CZ, SI, EE, EL

**ELIXIR ESIF Task Force (Months 1-12)**

ELIXIR Structural Funds Task Force grouping funding specialists across ELIXIR Nodes will be established to share best practice in ESIF use for research infrastructures. The Task Force would also engage external experts such as ones from national managing authorities for ESIF, DG REGIO, DG EMPLOY, DG Enterprise and Industry and Jaspers and would make use of existing reports such as the ESPON KIT report (www.espon.eu).

An ELIXIR-wide Workshop early at start of the project to pool good practice on using Structural Funds to support research infrastructures and facilitate personal interactions. Meeting will be hosted and organised by CEITEC, who leads this task.

This would include talks from ELIXIR Nodes with experience of accessing Structural Funds (Estonia, Czech Rep, Slovenia), as well as other ESFRIs such as ELI that have done this successfully in other disciplines

Local priorities and their overlaps identification towards Business Case (Months 6-24)

As all regional priorities are different, and as the application process for funding is done in the local language and following local rules, target Nodes will work with their regional partners to understand the priorities. This task will support Nodes in understanding their local Smart Specialisation Strategy and the regional priorities relating to research and life sciences. Access support from Jaspers following the connections built up within Months 1-12.

Supporting Nodes in actually developing the Business Cases and applications for Structural funds to support the construction and/or operation of the Node. The timing of this work will depend on when the calls will be opened for each region.

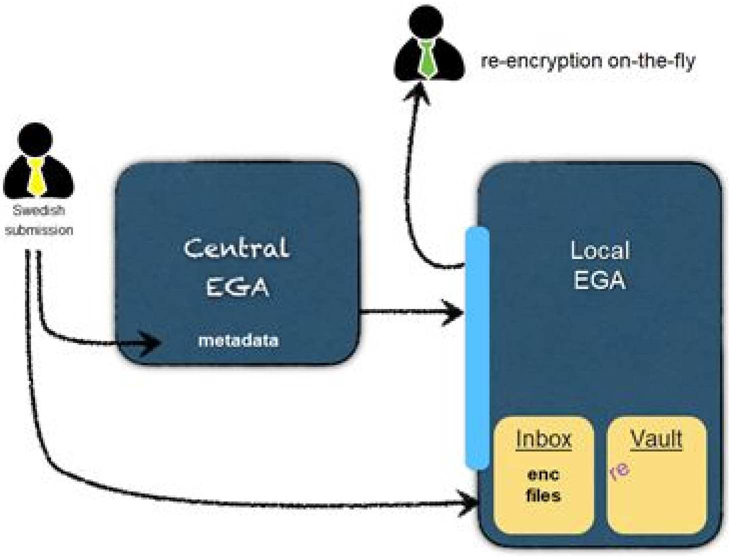
Partners: CZ, SI, EE, EL

# 7. Appendix 1: How to set up a local EGA node

## 7.1. Introduction

Human biomedical research data is often sensitive in nature, and projects have strong requirements on IT security, data access control, and data management to protect the privacy of individuals, the integrity of research and the adherence of legal or contractual obligations. This is particularly true for collaborations between academic and non-academic partners from industry or healthcare. As a consequence of this the European Genome-phenome Archive (EGA) was established at EMBL-EBI in 2008, and later on the Centre for Genomic Regulation (CRG) (Spain) joined in the operation. Hitherto, the EGA has been operated as a centralised structure hosted by full mirror sites at the EBI and CRG.

The rapid increase in human genome sequence data, and the heterogeneous legal landscape among different countries for this type of data, makes it evident that a centralised structure would not be a scalable solution for the future. Therefore, EGA and ELIXIR have identified the need to create a federated EGA organisation. In this setup, local data archiving instances can be established that interfaces with the central EGA to create a distributed but still coherent infrastructure for storing and distributing human genome sequence data. Work to make this federated setup a reality has been carried out within ELIXIR-EXCELERATE WP9 and WP10. The Local EGA efforts aim at producing portable code that can be deployed at sites that want to be able to archive sensitive human genetic data locally, while still being discoverable and accessible through the central EGA interfaces, for researchers that have legitimate reasons for accessing the sensitive data.



The basic functionality of local EGA is shown in Figure 1. A submitter, that wants to deposit data at a local instance, will indicate through the user interfaces at Central EGA that data will be deposited at a particular local instance, and then upload the data files to the local instance using the same user credentials as at Central EGA. Dataset submission will then be done by supplying the necessary metadata through the Central EGA user interface, which will signal to the local instance that the data files shall be archived (re-encrypted and moved into a secure vault storage area). Datasets are discoverable through the Central EGA interfaces, where data access requests are processed. Datasets for which access has been granted are retrievable from the local instance using the EGA user credentials, and is re-encrypted for the particular user at the time of access.

A local EGA node will be instrumental for each ELIXIR node in storing and making available sensitive research data produced in the respective country. A local EGA node will enable re-use of human genetic and phenotypic data in a protected environment that only allows legally correct access in accordance with ethical permits of the respective scientific studies. Thus, the local EGA node will make genetic data FAIR (Findable, Accessible, Interoperable and Reusable).

The local/federated EGA forms a sustainable, secure and legally correct storage of sensitive human genetic data, which facilitates data publication and open science, which in turn enables data sharing of benefit for scientists, healthcare providers and industry. Furthermore, access to European data (and in the future global data) accelerates life sciences and paves the ground for new directions of research. This access to human data is of special importance for healthcare providers when treating patients with rare diseases having genetic variants that are only known in a few cases world-wide.

In this report we describe how to set-up a local EGA node in the federated landscape of EGA nodes.

## 7.2. Systems development for local EGA

Systems development for local EGA has engaged staff across multiple ELIXIR nodes, including Spain, Sweden, Finland, Norway, and EBI. Funding has been provided through ELIXIR-EXCELERATE, as well as the Nordic Tryggve project funded by the Nordic e-Infrastructure Collaboration (NeIC). Regular teleconference meetings have been held every second week for coordination of the project.

The aim of the joint development activities has been to establish a code-base that supports the core functionality of

1. Data In (by a Submitter)
    - Submitter inbox creation
    - Data upload and archiving
2. Data Out (by an authorised Requestor)

A video recording demonstrating these functionalities is available on YouTube[2].

The source code is available from these GitHub repositories:

- Submission part: https://github.com/EGA-archive/LocalEGA

[2] https://www.youtube.com/watch?v=d8tIDZvDGKQ

- Data access part: https://github.com/EGA-archive/ega-data-api
- Encryption tool: https://github.com/EGA-archive/crypt4gh

## 7.3. Overview of how a local EGA node can be established

The Local EGA instances will be integrated with the current EGA mirrors in such a way that submissions, which are done by specifying community standard metadata for studies, samples and analyses, will be managed through the central interfaces at CRG or EBI, whereas the actual datasets will be housed at local instances. This will make datasets globally findable through the central interfaces, regardless of where the datasets are stored physically.

### 7.3.1. System architecture

The system in the first instance relies on user interfaces at the Central EGA for: Submitter registration, Metadata submission, and Requestor authorization, using messaging mechanisms between the Central EGA and a local instance through securely connected RabbitMQ queues. This enables the establishment of a federated EGA set-up where there is a unified interface for submission, querying and data access requests.

The system is built on containerized microservices. The *Data In* components have mainly been developed in the activities of the Tryggve project, whereas the *Data Out* builds on the EGA Data API v3 originally developed by EGA staff. As Data In and Data Out were developed independently of each other, there were microservices with overlapping functionality in both. This overlap has now been removed by the cross-organisational development team to produce code that enables integration of the two.

#### 7.3.1.1. *Access control*

Data submitters and requestors will be provided with accounts through the already existing account management procedures at the Central EGA, and a mapping of these authentication and authorization credentials to the ELIXIR AAI federated identity management system is underway.

#### 7.3.1.2. *Data In*

The system allows a Submitter registered at Central EGA to access an inbox at a local EGA instance, and upload files to this inbox. The inbox area will be created on the fly at the local instance with credentials from the Central EGA. Uploaded files are encrypted in the crypt4gh format[3] before uploading. This format makes it possible to change the cryptographic protection of the whole file by re-encrypting a minor part it, which makes it

[3] https://docs.google.com/presentation/d/1Jg0cUCLBO7ctyIWiyTmxb5Il_fQVzKzrxHHzR0K9ZvU/edit#slide=id.g3b7e5ab607_0_2

possible to re-encrypt the file with different keys for different recipients in a very compute efficient way.

The submitter then attaches the necessary metadata to the dataset files, and triggers the archiving of the files uploaded to the local instance inbox, through the Central EGA submission interfaces. Relevant parts of the uploaded files are then re-encrypted in a way that makes it specific for the particular instance, and the files are moved from the inbox to a protected archive storage area.

#### 7.3.1.3. *Data Out*

A Requester that has been authorized to access a particular dataset is able to receive the content of the files in this dataset by accessing a link along with an authorisation token that contains the file credentials. For supported file formats the link provides an htsget protocol[4] encrypted data stream, which allows retrieval of whole or parts of a genome sequence file. Upon access, the file will be re-encrypted with a key that is unique for the recipient.

#### 7.3.1.4. *Microservice separation*

The architecture of the system is set-up such that only few of the microservice components are exposed to the outside world. Basically, the only outward-facing parts are the inbox component which does not have access to the rest of the system, and the file access endpoint that will only respond to valid requests with authorized access tokens. There are a couple of other components that have outgoing connections for messaging, and these are through encrypted channels protected by certificates.

#### 7.3.1.5. *Deployment*

Deployment strategies of the containerized microservices have been developed for Kubernetes and Docker Swarm. The whole code base has been successfully deployed in Kubernetes environments in Finland and Sweden.

More in-depth documentation of the software components and how to deploy them is available at on the Local EGA[5].

### 7.3.2. Legal

Given that it is most likely that the majority of Local EGA nodes that will be established by legal entities located in the EU (at least initially), this section focuses on the current legislative situation in the EU. Data deposited in EGA is legally defined as personal data under the EU General Data Protection Regulation (Regulation (EU) 2016/679; “GDPR”), when the data is derived from individuals in the EU. As it relates to genetic information and/or health, much of it is also considered to be belonging to the special categories of

[4] http://samtools.github.io/hts-specs/htsget.html

[5] https://localega.readthedocs.io/en/latest/

personal data (often called “sensitive” data) for which stricter rules apply. The entity who decides on why and how personal data should be processed is called **Controller**. A Controller can decide to use another entity to help process the data. That entity is called a **Processor**. As the governance process of granting access to data deposited in the EGA is solely in the hands of the Data Access Committee (DAC) entity defined by the data depositor, the legal entity that will host the local EGA instance will be a *de facto* Processor as defined by the GDPR. Article 28 of the GDPR the GDPR clearly states that a legally binding agreement must be established between the Controller and the Processor. Article 28 also outlines the mandatory content of such an agreement. Note that the Controller and Processor roles, and hence the legal responsibilities, in most cases lie with the legal entities that employ the researchers that deposit the personal data, rather than with the individual researchers. The Local EGA hosting party will basically have three options for ensuring that the legal agreement requirements are covered when hosting the deposited data:

- Establish separate Data Processing Agreements between the Local EGA hosting entity and the depositing entity, for each deposited dataset.
- If the numbers of depositing entities are limited, general agreements that cover all datasets submitted from a depositing entity might be established.
- A hybrid approach, where a general agreement is done between the parties, and then an amendment is agreed on for each separate dataset.

As legislation in this area varies in some respects in different countries, the Local EGA hosting entity should consult the Data Protection Officer of the institution, and the national Data Protection Authority of the country in question, when deciding on how to handle these issues.

As stated above, the legal responsibility regarding deciding if, and under what conditions, datasets can be shared lies solely with the depositor as Controller. It is likely that in most cases the receiving legal entity will become a Controller of the personal data as well. The Controller should make sure that the terms and conditions for use of the data by the receiving party is clearly regulated in a legal agreement, even though it is not stated in the GDPR how this should be done. Indeed the EGA requires the depositor of a dataset to define a Data Access Agreement (DAA) for each dataset. A Local EGA hosting entity could facilitate data depositions by drafting template DAAs that are tailored to the local legal environment. However, the Local EGA hosting entity should make it clear to the depositor that any legal responsibilities when it comes to data sharing lies with the depositor alone.

Apart from the GDPR, there are other pieces of legislation that will affect how and if data can be shared with others, e.g. regarding ethical review and informed consents. In this area there is no pan-EU legislation, so the situation will vary depending on the local national legislation. However, the responsibility for this lies solely with the data depositor, and should be made aware that the Local EGA hosting entity will not accept any responsibility regarding these issues.

### 7.3.3. Agreements central EGA -- local EGA

EGA has circulated a proposal for comments for the organization of a Federated EGA in March 2019. It is envisioned that it will be structured according to a tiered model, with two different levels for the local instances.

Level 1 nodes will enter into a legal agreement with Central EGA for a (suggested) minimum membership term of four years, and shall offer “full” EGA services - supporting the full data deposition process through the standard EGA submission user interfaces (or a separate interfaces), worldwide data distribution, metadata sharing, and providing a local helpdesk function. They will integrate authentication and authorization infrastructure with the Central EGA. Level 1 nodes are expected to participate in the development of common EGA APIs, tools, and resources. They shall be represented in strategic and operational governing committees.

Level 2 nodes are expected to be various kinds of human data distribution hubs, such as e.g. individual institutions and research consortia. These are not expected to accept external data submissions, and handle data access according to their defined processes, but exchange shareable metadata with the EGA for discoverability. Level 2 nodes will enter into a MoU with EGA for a (suggested) minimum membership term of four years, and will be observers in strategic and operational governing committees.

Draft legal agreements and/or MoUs with the first nodes are to be presented to the first local nodes before end of August 2019.

## 7.4. Conclusion

Here we have provided the basis for setting up a local EGA data node within ELIXIR. As time of writing, three such nodes have been set up during the framework of ELIXIR-EXCELERATE -- Sweden, Finland and Norway. Further nodes are expected within the next years.