

# Deliverable D7.4 Report on operational processes across the federated EGA network and data transfer with the European COVID-19 platform

Project Title (Grant agreement no.):	ELIXIR-CONVERGE: Connect and align ELIXIR Nodes to deliver sustainable FAIR life-science data management services (871075)		
Project Acronym (EC Call):	ELIXIR-CONVERGE (H2020-INFRADEV-2018-2020)		
WP No & Title:	WP7 Federated European Genome-phenome Archives for transnational access of COVID-19 host data		
WP leader(s):	Jordi Rambla (CRG) & Thomas Keane (EMBL-EBI)		
Deliverable Beneficiary:	Lead	10 - BSC	
Contractual delivery date:	30/09/2022	Actual delivery date:	02/11/2022
Delayed:	No		
Partner(s) contributing to this deliverable:	BSC, EMBL-EBI, CRG		
Authors: Albert Hornos Vidal (BSC), Laura Portell Silva (BSC), Salvador Capella-Gutierrez (BSC), Mallory Freeberg (EMBL-EBI), Amy Curwin (CRG)			
Contributors: Jordi Rambla (CRG)			
Acknowledgments (not grant participants): -			
Reviewers:	ELIXIR-CONVERGE Management Board (MB) members.		

## Log of changes

DATE	Mvm	Who	Description
30/09/2021	v1	Salvador Capella-Gutierrez (BSC)	Initial version
30/09/2022	v2	Albert Hornos Vidal (BSC) Laura Portell Silva (BSC) Salvador Capella-Gutierrez (BSC) Mallory Freeberg (EMBL-EBI) Amy Curwin (CRG)	Final version pending feedback by reviewers
18/10/2022	v3	Albert Hornos Vidal (BSC) Laura Portell Silva (BSC)	Sent to PMU after incorporating internal WP feedback

		Salvador Capella-Gutierrez (BSC) Mallory Freeberg (EMBL-EBI) Amy Curwin (CRG)	
19/10/2022	v4	Nikki Coutts (ELIXIR Hub)	Circulated to the MB for final review before submission
01/11/2022	v5	Salvador Capella-Gutierrez (BSC)	MB comments addressed
02/11/2022	1v0	Nikki Coutts (ELIXIR Hub)	Final version to be uploaded into EC Portal

## Table of contents

<b>1. Executive Summary</b>	<b>3</b>
<b>2. Contribution toward project objectives</b>	<b>4</b>
<b>3. Introduction</b>	<b>7</b>
3.1 Scope of Deliverable	7
3.2 Relationship with other WPs	7
3.3 Methodology	7
<b>4. Description of work accomplished</b>	<b>7</b>
4.1 Federated EGA Maturity Model	8
4.1.1 Background on maturity models.	8
4.1.2 Other existing maturity models	8
4.1.3 Federated EGA Maturity Model: vision, mission and development.	8
4.1.4 Dissemination activities	9
4.2 Federated EGA Knowledge Base	9
4.3 Standard Operating Procedures (SOPs)	9
4.4 Intersecting EGA and the COVID-19 Data portal	10
<b>5. Results</b>	<b>12</b>
5.1 Federated EGA Maturity Model	12
5.1.1 Federated EGA Maturity Model	12
5.1.1.1 Domains	12
5.1.2 Federated EGA Maturity Model Self-assessment tool	13
5.1.3 Aggregated results from the initial evaluations.	14
5.2 Federated EGA Knowledge Base	16
5.3 Standard Operating Procedures (SOPs)	18
5.3.1 User-facing Processes	18

5.3.2 Internal Node Processes	18
5.3.3 Joint CEGA/FEGA Node Processes	19
<b>6. Conclusions</b>	<b>20</b>
<b>7. Impact</b>	<b>21</b>
<b>8. Next Steps</b>	<b>21</b>
<b>9. Deviation from Description of Action</b>	<b>21</b>



## 1. Executive Summary

ELIXIR Converge WP7 was conceived as a mechanism to accelerate the creation of the Federated version of the European Genome-phenome Archive (FEGA) in the context of the COVID-19 pandemic. As part of this effort, a Node maturity model has been designed, deployed and tested across the Nodes of the federation. The Node maturity model aims to become a self-guided tool covering most of the relevant aspects to be taken into account when deploying and operating a federated EGA Node. As part of a concerted effort to manage and share COVID-19 host genomics data, this deliverable illustrates the connection of the development of the federated EGA and its alignment with the national and regional plans for data-sharing initiatives.

Since ELIXIR Converge WP7 started, the work of WP has been focused on understanding the needs for a structured model to facilitate the engagement of the new Nodes to the Federated EGA Network. For this reason, several resources have been created to help new Nodes to start the work.

The main activities described in this deliverable are related to the FEGA Maturity Model and operational documentation for FEGA Nodes, including the main resources that can be consulted to facilitate the operation of a FEGA Node. To centralise all these resources, the Federated EGA Knowledge Base was assembled following the Maturity Model structure with the information on how to join the Federated EGA Network based on the knowledge and experiences of current Nodes and their use cases. Indeed, the FEGA Maturity Model and FEGA Knowledge Base provide information across the relevant domains: Governance, Strategy & Sustainability, Legal, Data & Metadata management, Technical Infrastructure, Operations support and Communications, Community Building, and Engagement. Each of these domains emphasise important aspects to consider when establishing a FEGA Node.

The ultimate goal of this deliverable is to provide future and existing FEGA Nodes with resources on how to follow best practices when implementing the Node. In the same way, it intends to agree on some guidelines to standardise the processes among FEGA Nodes as much as possible and, thus, facilitate the transit among them as well as the transit between the central Node of EGA and the FEGA Nodes. Importantly, these outcomes are instrumental to the European Genomics Data Infrastructure (GDI) as FEGA Nodes are a core component of this project. Indeed, the Federated EGA Maturity Model and Knowledge Base will serve as important elements for having 15+ operational national GDI nodes by 2026.

## 2. Contribution toward project objectives

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

Objective no. / Key Result no. Description	Contributed to:
<b>Objective 1:</b> Develop a sustainable and scalable operating model for transnational life-science data management support by leveraging national capabilities ( <b>WP1, WP5</b> )	
<b>Key Result 1.1:</b> Established European expert network of data stewards that connect national data centres and similar infrastructures and drive the development of interoperable solutions following international best practice, including national interpretations of the General Data Protection Regulation (GDPR)	Yes
<b>Key Result 1.2:</b> Development of joint guidelines and common toolkit that are adopted into funder recommendations, with support available nationally and in local languages	Yes
<b>Key Result 1.3:</b> The catalogue of successful national business models incorporated into national strategies	Yes
<b>Key Result 1.4:</b> The developed “sustainable and scalable operating model for transnational life-science data management support” is adopted into national ELIXIR Node	Yes
<b>Objective 2:</b> Strengthen Europe’s data management capacity through a comprehensive training programme delivered throughout the European Research Area ( <b>WP2, WP6</b> )	
<b>Key Result 2.1:</b> A comprehensive ELIXIR Training and Capacity building programme in Data Management, directed at both data managers and ELIXIR users, and connected to the national training programmes in Data Management in the ELIXIR Nodes and prospective ELIXIR Member countries.	Yes
<b>Key Result 2.2:</b> Development of a collective group of trainers that support scalable deployment of Data Management training across ELIXIR Nodes.	No
<b>Key Result 2.3:</b> A substantial cohort of data managers, Node coordinators and researchers with specific data management skills, business planning and knowledge of transnational operations across the ELIXIR Nodes	Yes
<b>Objective 3:</b> Align national data management standards and services through a sustainable, scalable and cost-effective data management toolkit ( <b>WP2, WP3, WP5</b> )	

<b>Key Result 3.1:</b> Assemble a full-stack harmonised common toolkit comprising all aspects of data management: from data capture, annotation, and sharing; to integration with analysis platforms and making the data publicly available according to international standards.	<b>Yes</b>
<b>Key Result 3.2:</b> Provide exemplar toolkit configurations for prioritised demonstrators to serve as templates for future use.	<b>No</b>
<b>Key Result 3.3:</b> Establish national capacity in using as well as updating, extending and sustaining the toolkit across the ERA.	<b>No</b>
<b>Key Result 3.4:</b> Enable 'FAIR at source' practice for data generation, and analytical process pipeline implementation by flexible deployment of the toolkit in national operations	<b>No</b>
<b>Objective 4:</b> Align national investments to drive local impact and global influence of ELIXIR (WP4,WP6)	
<b>Key Result 4.1:</b> Development of a Node Impact Assessment Toolkit based on RI-PATHS methodology.	<b>No</b>
<b>Key Result 4.2:</b> Adoption of Impact assessment in ELIXIR Nodes, supported by Node coordinators network and feedback on applicability from dialogues with national funders.	<b>No</b>
<b>Key Result 4.3:</b> Creation of national public-private partnerships and industry outreach where open life-science data and services stimulate local bioeconomy	<b>No</b>
<b>Key Result 4.4:</b> Growth in reach, impact and engagement of stakeholder communication assessed by established ELIXIR Communications metrics	<b>No</b>
<b>Key Result 4.5:</b> Initiating and advancing discussions on Membership (EU and international) or strategic partnerships (international countries) following ELIXIR-CONVERGE workshops.	<b>Yes</b>
<b>Objectives - WP7 - Federated European Genome-phenome Archives for transnational access of COVID-19 host data.</b>	
<b>O7.1:</b> Federation architecture, interfaces, and compliance tests for the Federated EGA network (Task 7.1).	<b>Yes</b>
<b>O7.2:</b> Coordination of the development of a reference implementation sufficient to create a functional Federated EGA Node (Task 7.1).	<b>Yes</b>

<b>07.3:</b> Development of phenotype metadata model to enable mapping and linkage of COVID-19 host-clinical measures across European national Nodes, and robust linkage between host and viral datasets (Task 7.3).	<b>No</b>
<b>07.4:</b> The development of documentation and guidelines for the operational practices of federated EGA Nodes (Task 7.4).	<b>Yes</b>

## 3. Introduction

### 3.1 Scope of Deliverable

This deliverable includes the description and links to documentation, best practices, and SOPs for operating the Federated EGA Network Nodes. In the same way, it details a central element on which the rest of the documentation has been based, the [Federated EGA Maturity Model](#)<sup>1</sup>. The FEGA Maturity Model has served as reference to structure the rest of the documentation developed for the establishment and maintenance of a FEGA Node.

The main activities described in this deliverable are related to the FEGA Maturity Model and the [FEGA Knowledge Base](#)<sup>2</sup>. The FEGA Maturity Model is a tool to assess how good and capable of continuous improvement a Node of the FEGA Network is. In the same way, the FEGA Knowledge Base is a repository, in web format, of useful documentation when establishing a FEGA Node.

### 3.2 Relationship with other WPs

- ELIXIR Converge WP2 members contributed to the creation of training materials documentation of the FEGA Knowledge Base.
- ELIXIR Converge WP7 members contributed to the WP9 September 2022 "Brokering Maturity Model" Workshop in Padua
- ELIXIR Converge WP7 members attended the WP3 contentathons to contribute to the content of the '[Human data domain page](#)<sup>3</sup>' of the RDMkit.

### 3.3 Methodology

To complete the work for this deliverable, a working group was established which included representatives of various Nodes of the FEGA Network, representatives of the Central EGA Node and a representative from the ELIXIR Hub with ample experience on developing maturity models across different ELIXIR projects.

For approximately one and a half years, April 2021 to June 2022, biweekly meetings were held to agree on the domains, subdomains, and indicators resulting from the Maturity Model. For the development of the FEGA Knowledge Base and to organise the different SOPs available, a smaller

<sup>1</sup> <https://inab.github.io/feqa-mm/>

<sup>2</sup> <https://ega-archive.github.io/FEQA-onboarding/>

<sup>3</sup> [https://rdmkit.elixir-europe.org/human\\_data](https://rdmkit.elixir-europe.org/human_data)

group was set up to carry out specific work sessions during June, July and September 2022.

## 4. Description of work accomplished

The different activities done for this deliverable are focused on creating guidelines and best practices for new and existing FEGA Network Nodes. This effort will provide potential network Nodes with reference documentation in specific domains, which can assist them in the implementation and deployment of a FEGA Network Node as well as in the periodic evaluation of their performance against a defined set of indicators. The six different domains are divided into Governance, Strategy & Sustainability, Legal, Data & Metadata management, Technical Infrastructure, Operations support and Communications, Community Building, and Engagement. Those domains are subsequently divided in a total of 19 subdomains, which comprise 36 different indicators.

### 4.1 Federated EGA Maturity Model

#### 4.1.1 Background on maturity models.

Maturity models represent conceptual tools designed to advance in the implementation, deployment and maintenance of complex systems. They initially were used to measure capabilities in software development and help determine weak points. Thus, the same principles used for software development's maturity models were followed to develop the FEGA Maturity Model, which can serve as a tool to establish new Nodes and to identify areas in need of improvement.

#### 4.1.2 Other existing maturity models.

The FEGA Maturity Model was developed following similar efforts across other ELIXIR projects. The objective of this coordinated work is to ensure global consistency across the way maturity models are structured, e.g. into domains, subdomains and indicators; the scale for levels associated to indicators, from 1 to 5, and the overall meaning of levels, with level 1 focus on raising awareness about a particular aspect and level 5 representing the periodic evaluation of a specific item.

Two maturity models have been important for the development of the FEGA Maturity Model:

- [B1MG Maturity Level Model Framework](https://b1mg-project.github.io/MLM/)<sup>4</sup>, which focuses on the different capabilities that a national healthcare system would need to support genomic medicine, with connections into research.
- [The ELIXIR Human Data Infrastructure Maturity Model](https://elixir-europe.github.io/human-data-maturity-model/)<sup>5</sup>, which is designed to give Nodes a tool to guide advancement within the European network for human data sharing in a harmonised fashion.
- [FAIRplus Dataset Maturity Model](https://fairplus.github.io/Data-Maturity/)<sup>6</sup>, which focuses on the iterative processes that a given dataset should follow to increase its FAIRness level. Despite being centred around datasets, level 5 for this maturity model implies an organisational change to ensure that any dataset produced in a given organisation is FAIR by design.

<sup>4</sup> <https://b1mg-project.github.io/MLM/>

<sup>5</sup> <https://elixir-europe.github.io/human-data-maturity-model/>

<sup>6</sup> <https://fairplus.github.io/Data-Maturity/>



#### 4.1.3 Federated EGA Maturity Model: vision, mission and development.

The FEAGA Maturity Model is envisioned as a mechanism to drive engagement and disengagement of Nodes being part of the FEAGA. Following existing maturity models across other ELIXIR projects, the initial work revolved around proposing a set of consistent domains, subdomains and associated indicators needed when conceptualising, developing, deploying and maintaining a FEAGA Network Node. The initial proposal was iteratively revised to ensure consistency and that it is fit-for-purpose. Then, indicators were further developed assigning them different levels from 1 to 5. Level 1 was directed towards initial awareness on specific aspects, levels 2 and 3 represent actual work carried out to increase the maturity of the indicator, and level 4 represents a fully operational mode. Level 5 is designed for the periodic review of the indicator and the adoption of the most up-to-date technical and legal specifications. Indicators were grouped into three main categories (Essential, Important and Useful) depending on their relevance towards establishing a fully operational FEAGA Node. Essential indicators (seven indicators, representing 19% of the total indicators) were required to have a minimum level of 4. Important indicators (14 indicators, representing 39% of the total) were deemed to have a minimum level of 3. Having those expected levels for this set of indicators (21 indicators out of 36 spanning the six proposed domains) allow identification of FEAGA Nodes that are fully operational and can enter shortly in production. Useful indicators (15) represent aspects that should be taken into account - especially for the sustainability of the Node - but are not considered necessary for a Node to enter into production. Useful indicators are required to have a minimum level of 2, representing that the Node is aware of those requirements and is working towards higher levels.

The initial version of the FEAGA Maturity Model was evaluated by the inaugural FEAGA Network Nodes, the Central EGA, and a new institution willing to set up a FEAGA Node in their jurisdiction. The plan was to evaluate the consistency and specificity of the proposed structure divided in domains, subdomains and indicators, as well as to gather further feedback on the usefulness of the domains. Importantly, this exercise served to identify areas that needed additional clarification as different FEAGA Nodes understood them differently. Two important aspects were also identified during this process: 1) it is important to provide examples per indicator to facilitate self-assessment, and 2) it is important to clarify what computational needs are, regarding a FEAGA Node. As hosted data is usually accessed for further analysis, the indicator about computational needs might be understood as having enough computational power for being able to provide analysis capabilities to authorised users. However, this indicator refers to the computational capabilities needed by a FEAGA Node for day-to-day operations, e.g. quality controls over submissions, encryption, decryption and re-encryption of hosted data.

#### 4.1.4 Dissemination activities

The Maturity Model was presented at the 8th ELIXIR All Hands meeting that took place from June 7th to 10th, 2022, in a workshop entitled: **Federated European Genome-phenome Archive (FEAGA): state of the art, Q&A and demonstration sessions**. In this workshop, the Maturity Model was presented to the broad audience, including the aggregated results obtained after the individual evaluation of the Maturity Model by the five inaugural FEAGA Network Nodes and Central EGA. Relevant feedback obtained during this workshop will be incorporated into future iterations of the Maturity Model. This public presentation completed a number of dissemination activities carried out

to promote the development and attract initial feedback around the development of the FEGA Maturity Model.

## 4.2 Federated EGA Knowledge Base

Once the first version of the Maturity Model was released, another working group was created in collaboration with the Federated EGA Operations Committee. This working group was tasked with the development of the Federated EGA Knowledge Base. The FEGA Knowledge Base, organised as a collaborative web-based space, offers materials for establishing a Node in the Federated EGA Network. The materials are based on current Nodes' knowledge and experiences, and use cases, and they will be expanded as the different Nodes contribute new use cases in the different sections. The FEGA Knowledge Base is deemed as the perfect complement to the FEGA Maturity Model as it includes real-life references for the establishment of a Node, as well as other training materials relevant to the different stakeholders involved in setting up a FEGA Node.

Interestingly, the initial version of the FEGA Knowledge Base was established during a two-day hackathon held in Barcelona in July 2022, and further developed over the following weeks until reaching its current structure. We envision the FEGA Knowledge Base to be a community-maintained resource that evolves over time with input from FEGA community members. To this end, the GitHub repository hosting the Knowledge Base is public, and a Contributing Guide has been drafted to facilitate community contributions.

## 4.3 Standard Operating Procedures (SOPs)

Standard Operating Procedures (SOPs) are important guides for the operation of any service. This is especially important in the context of the Federated EGA Network considering the always evolving technological landscape and the periodic review of the legal framework where FEGA Nodes operate. Indeed, the possible expansion of the FEGA Network beyond Europe to countries like Canada, Australia and Argentina will have a direct impact on how the Nodes carry on their operations, especially the ones requiring exchanging metadata with Central EGA and, potentially, other Nodes in the network.

In this framework, and in collaboration with the FEGA Operations Committee, three sets of SOPs have been provided to the FEGA Nodes. These SOPs are adaptations of the existing ones from the Central Node for operating across a collaborative network of Nodes. It is expected that FEGA Nodes will take those SOPs and adapt them to their particular set-up. Feedback on these SOPs will drive their eventual evolution over time and the possibility of making them more granular.

These three sets of SOPs revolve around:

- **User-facing Processes**, which focuses on the iterations between a FEGA Node and users willing to make a submission to EGA, as well as those requesting access to a particular dataset. It also elaborates on the steps to follow for unresponsive Data Access Committees, which are responsible for granting access to datasets under their control.
- **Internal Node Processes**, which highlights how a FEGA Node can organise itself, e.g. using a ticketing system, to provide a timely and consistent experience to different stakeholders.

- **Joint CEGA/FEGA Node Processes**, which establish the interactions between a FEGA Node and the Central Node when it comes to the submission of any dataset to the Node. It is important to highlight that these SOPs are mandatory to ensure the correct management of metadata associated with any submitted dataset, as well as to the management of users submitting data/metadata.

It is important to emphasise that these SOPs must be understood as dynamic and configurable documents. The final definition of each SOP will depend on the specific needs of each Federated EGA Network Node.

## 4.4 Intersecting EGA and the COVID-19 Data portal

The [COVID-19 Data portal](#)<sup>7</sup> was created as a joint effort between EMBL-EBI and ELIXIR, including its affiliated institutions. Key stakeholders joined in successive iterations of the data portal. The main objective of the COVID-19 Data portal is to facilitate the access to research data relevant to the study of the SARS-CoV-2 and COVID-19. EGA was chosen to host genomics data associated to COVID-19 patients that required access control, given the sensitive nature of this data. To scale-up the ingestion and distribution of this type of data, plans for building and deployment of Federated EGA were accelerated. To enable the possibility of managing host genomics data in both Central EGA and Federated EGA, the first objective was to understand relevant metadata associated with deposited data for enabling downstream applications. This objective was achieved as part of deliverable 7.3 on *COVID-19 metadata mapping model across COVID-19 studies in federated EGA*. Parallel to such efforts is how to expose that metadata through the COVID-19 Data Portal. EGA being a resource co-managed by EMBL-EBI facilitated the connection of EGA with the data portal. However, this effort only allowed to expose data available at Central EGA. It was decided to use this mechanism to expose metadata associated with data managed by FEGA Nodes. Such a decision was possible thanks to the joint management of metadata between Central EGA and each of the FEGA Nodes. New mechanisms have been designed and implemented to facilitate the metadata ingestion by the COVID-19 Data Portal in the framework of complementary projects, e.g. WP3 at BY-COVID. This project is part of a global effort to prepare the research community for future outbreaks.

## 5. Results

### 5.1 Federated EGA Maturity Model

#### 5.1.1 Federated EGA Maturity Model

As a result of the work done for this deliverable, a [web-page](#)<sup>8</sup> with the Maturity Model was published. This web-page should be used as a self-guide tool to conceptualise on the different aspects needed for setting-up and maintaining a Federated EGA Node.

As mentioned before, the FEGA Maturity Model comprises six domains, each of them containing one or more subdomains. Each of these subdomains contain a series of indicators that will be the ones

---

<sup>7</sup> <https://covid19dataportal.org>

<sup>8</sup> <https://inab.github.io/fega-mm/>

that will be evaluated on a scale of 1 to 5. Taking everything together, the FEGA Maturity Model is configured as follows:

- 6 domains
- 19 subdomains
- 36 indicators

#### 5.1.1.1 Domains

**[1] Governance, Strategy & Sustainability.** This domain contains all indicators related to the establishment, management and long-term operation of the Federated EGA Node. Indicators are deemed to provide a guideline to ensure strategic alignment with the Federated EGA ecosystem, while allowing each Node to respond to the requirement of their decision/policy/funding making bodies

**[2] Legal.** This domain contains all relevant aspects regarding the legal framework where Federated EGA Nodes operate

**[3] Data & metadata management.** This domain contains all indicators related to the secure management of sensitive data and associated metadata following existing norms and regulations at institutional, regional, national and, if appropriate, European levels.

**[4] Technical Infrastructure.** This domain contains all the indicators about the technical infrastructure needed for the FEGA Nodes to be part of the network. It includes software stack, technical benchmarking, storage system and network interfaces.

**[5] Operations Support.** This domain contains all indicators related to the Standard Operating Procedures (SOPs) and documentation needed for FEGA staff to complete their tasks in accordance with regulations and laws.

**[6] Communications, Community Building, and Engagement.** This domain covers all aspects related to the communication and outreach activities of the Federated EGA Node, including specific actions with relevant research communities. As part of the transversal activities, this domain also includes aspects on capacity building in connection with the Federated EGA ecosystem.

#### 5.1.2 Federated EGA Maturity Model Self-assessment tool

The [Federated EGA Maturity Model Assessment Tool](https://bit.ly/3ywuFgc)<sup>9</sup> is used for the self-assessment of a FEGA Node and to capture any supporting evidence of each independent indicator. The tool has been created as a guide to help apply the Maturity Model and, thus, help potential FEGA Nodes to identify both their strengths and weaknesses.

The assessment will be filled out by members of each Node. The members of the team will respond to one or more domains, according to expertise and the stakeholders' scope of activities. If a domain is covered by multiple members of the team, it can be filled out either a) individually, followed by a consensus meeting where the team decided on an agreed response, or b) as a group effort.

The domain experts should choose a Maturation Level Assessment based on the current status of the Node for each indicator. Optionally, the rationale for the choice of maturity level of each indicator can be provided.

---

<sup>9</sup> <https://bit.ly/3ywuFgc>



**Figure 1.** Screenshot for the Federated EGA Maturity Model web-page.

	Explanation	Subdomains	Indicators	Connected with	Maturity Levels	Indicator level type	Expected Maturation Level Goal	Maturation Lev Assessment
[1] Governance, Strategy & Sustainability	This domain contains all indicators related to the establishment, management and long-term operation of the Federated EGA node. Indicators are deemed to provide a guideline to ensure strategic alignment with the Federated EGA ecosystem while allowing each node to respond to the requirement of their decision/policy/funding making bodies	[1.1] Governance and structure of the Federated EGA node	[1.1.1] Dedicated governance bodies and structure defined for the Federated EGA instance	<div>1. Non existent.</div> <div>2. The team is assembled and proposed roles identified.</div> <div>3. Overall governance body and node structure is defined, with stakeholder consultation, and formally approved including key roles, e.g. DPO.</div> <div>4. Governance body is fully operating with key personnel and is monitored based on work plan.</div> <div>5. Governance body is institutionalized, protected from organizational changes, open to novel developments and supportive of international cooperation.</div>	Essential	4	1	
		[1.2] Vision and Strategy of the Federated EGA node	[1.2.1] Roadmap/plan defined for the Federated EGA instance	<div>1. No strategy/framework is defined.</div> <div>2. Initial strategy/framework being drafted.</div> <div>3. Strategy/framework approved.</div> <div>4. Strategy/framework being implemented with key personnel.</div> <div>5. Strategy/framework fully implemented with planning underway for the next 4 years.</div>	Useful	2	1	
		[1.3] Sustainability model of the Federated EGA node	[1.3.1] Immediate resources (short-to-mid term funding)	<div>1. No funding plan is available nor proposed.</div> <div>2. Initial costed plan being developed.</div> <div>3. Costed plan developed and approved.</div> <div>4. Costed plan developed implemented with some resources.</div> <div>5. Costed plan fully implemented with planning underway for the next 4 years.</div>	Essential	4	1	
			[1.3.2] Long-term sustainability	<div>1. No sustainability plan exists.</div> <div>2. Initial engagement with stakeholders to develop sustainability plan.</div> <div>3. Available funding identified and stakeholders have been mandated.</div> <div>4. Sustainability plan to support the FEGA is designed and being established. Other models for diversifying funding streams raising financial stakeholders are explored.</div> <div>5. Sustainable general plan is implemented including clear indications on funding periods and mechanisms for renewal.</div>	Important	3	1	
		[1.4] Overarching key performance indicators (KPIs)	[1.4.1] Implementation, adoption and usage of KPIs in the Federated EGA node	<div>1. No KPIs have been identified nor adopted.</div> <div>2. Set of KPIs for overall implementation of the federated EGA node work plan as well as for the global and individual records usage are drafted taking as reference the ones from the federated EGA network.</div> <div>3. Partial implementation of drafted KPIs.</div> <div>4. Full implementation of KPIs for the Federated EGA node.</div> <div>5. Sharing KPIs information across the federated EGA network as well as contributing to the periodic discussion of indicators being measured.</div>	Useful	2	1	

**Figure 2.** Screenshot of a partial view of the Federated EGA Maturity Model, including domains, sub-domains, indicators and the expected level for each of them. The last column includes the self-evaluation of an individual Node. By default, the level is assigned to the lowest possible value.



### 5.1.3 Aggregated results from the initial evaluations.

Initial evaluation of the FEGA Maturity Model was conducted by the five inaugural FEGA Network Nodes, the Central EGA and an additional institution that is exploring the possibilities of setting-up a FEGA Node in their jurisdiction. Every Node submitted their results anonymously and data was aggregated to have a general overview (Figure 3) of the initial evaluation. The initial results should serve two main purposes: evaluate the indicators' levels' consistency across Nodes when using similar evidence and detect unexpected patterns, e.g. low levels for a particular domain/subdomain. Unexpected patterns should help to prioritise efforts at the Federated EGA committees.

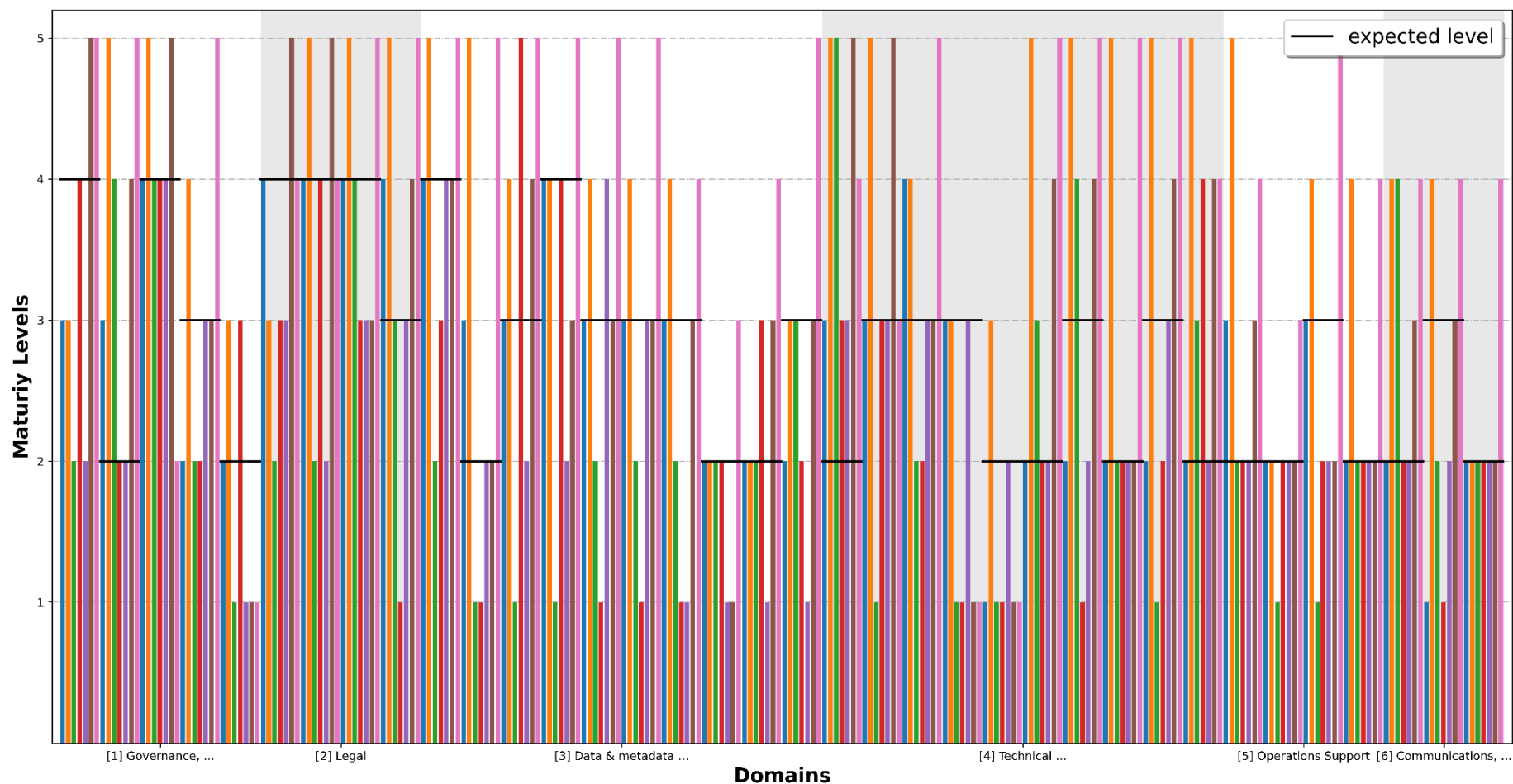
When analysing deeply the aggregated results obtained for the different FEGA Nodes, including the candidate one, and discarding Central EGA to prevent biases, there are a number of items that require attention. When looking at aggregated values for the first (**Governance, Strategy & Sustainability**) and second (**Legal**) domains, there are two indicators that require further work by all FEGA Nodes. This additional work should be at general level by providing additional information about the scope of those indicators, and at Node level by re-evaluating those indicators and providing further evidence in case higher scores are deemed.

- [1.1.1] Dedicated governance bodies and structure defined for the Federated EGA instance.
- [2.1.1] Data Protection Impact Analysis (DPIA) performed.

Indicator 2.2.1 has improved since this evaluation was performed as the five inaugural FEGA Nodes have signed their respective legal agreement.

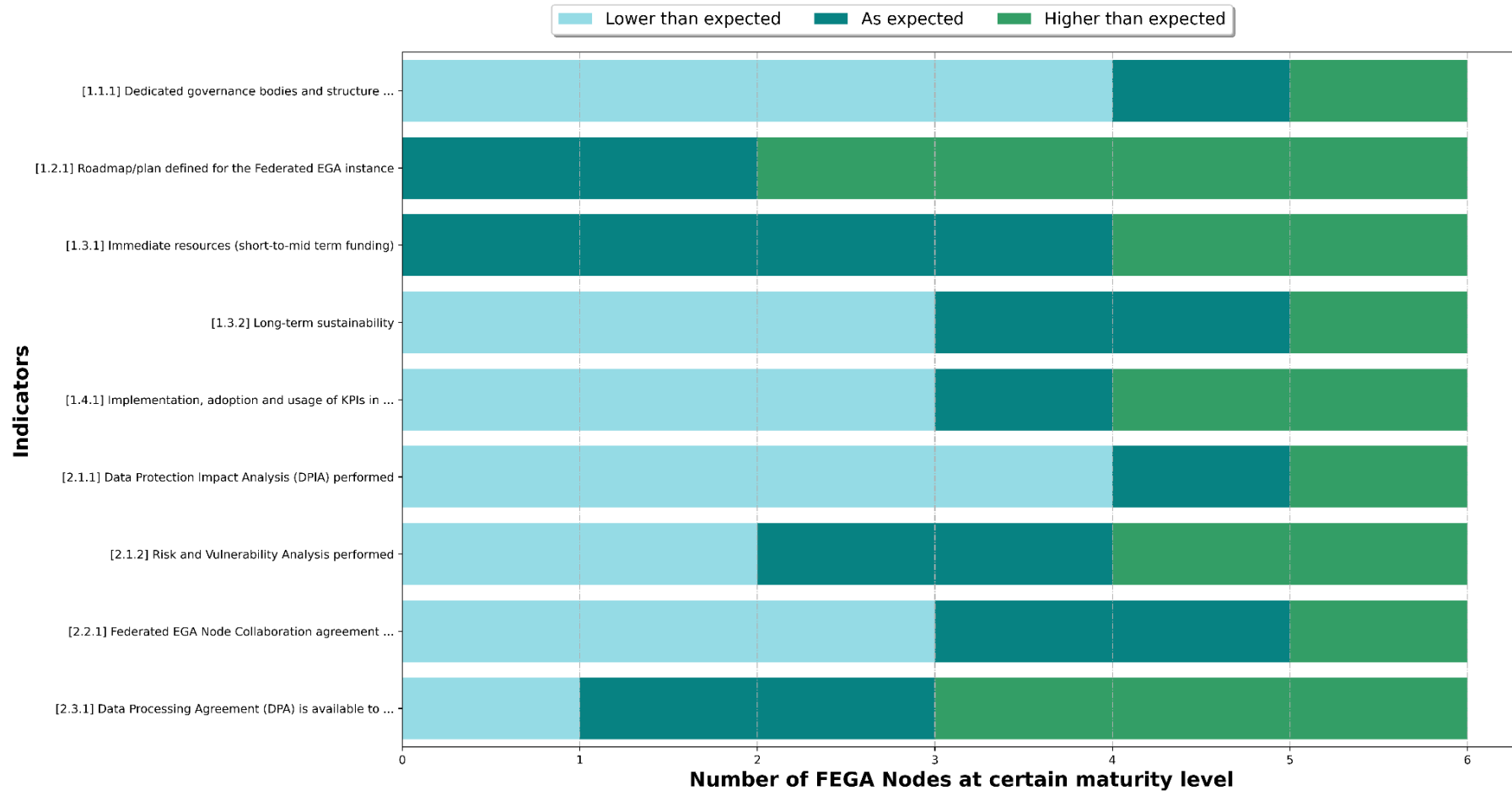
When analysing the technical domains, e.g. **Data & metadata management**, **Technical infrastructure**, there are interesting results to highlight. First, the adoption of Software Development Best practices by all FEGA Nodes when implementing, deploying and maintaining their sourcebase. Second, the need to better define the different tests to be performed by any FEGA Node regarding stress testing (4.2.2) and compliance testing (4.2.1). This requires a coordinated process by all Nodes at the FEGA Operations committee. Third, the five (out of the six considered Nodes) had already conducted their end-to-end technical demonstration of their deployed capacities. Therefore, it is not a surprise that associated indicators, e.g. 3.3.2, 4.1.2, to this demonstration show high level of compliance. Another interesting finding is the availability of computational power at each Node to perform all required operations, e.g. data ingestion and sharing, data encryption, decryption and re-encryption.

When analysing together the remaining two domains, **Operation Support** (5th), and **Communication, Community building and Engagement** (6th), we can also detect areas that need further improvements across the FEGA Nodes. The first indicator revolves around the FEGA Node **Helpdesk and SOPs** (5.1.3), which will consolidate as the Node enters in operation, and reference SOPs, provided as part of the FEGA Knowledge Base, are adapted to the Node. The second indicator that requires some attention is 6.2.1 on **Training for the Federated EGA Node users**, which refers to the training of what are known as *power users*, which refers to users that make an intensive use of the resource as submitters and/or data requesters and access to it. Power users that regularly access data in the FEGA Node might be interested in automating the process through programmatic APIs and the use of Secure Processing Environments (SPEs) - as close as possible to the data - for analysing it.



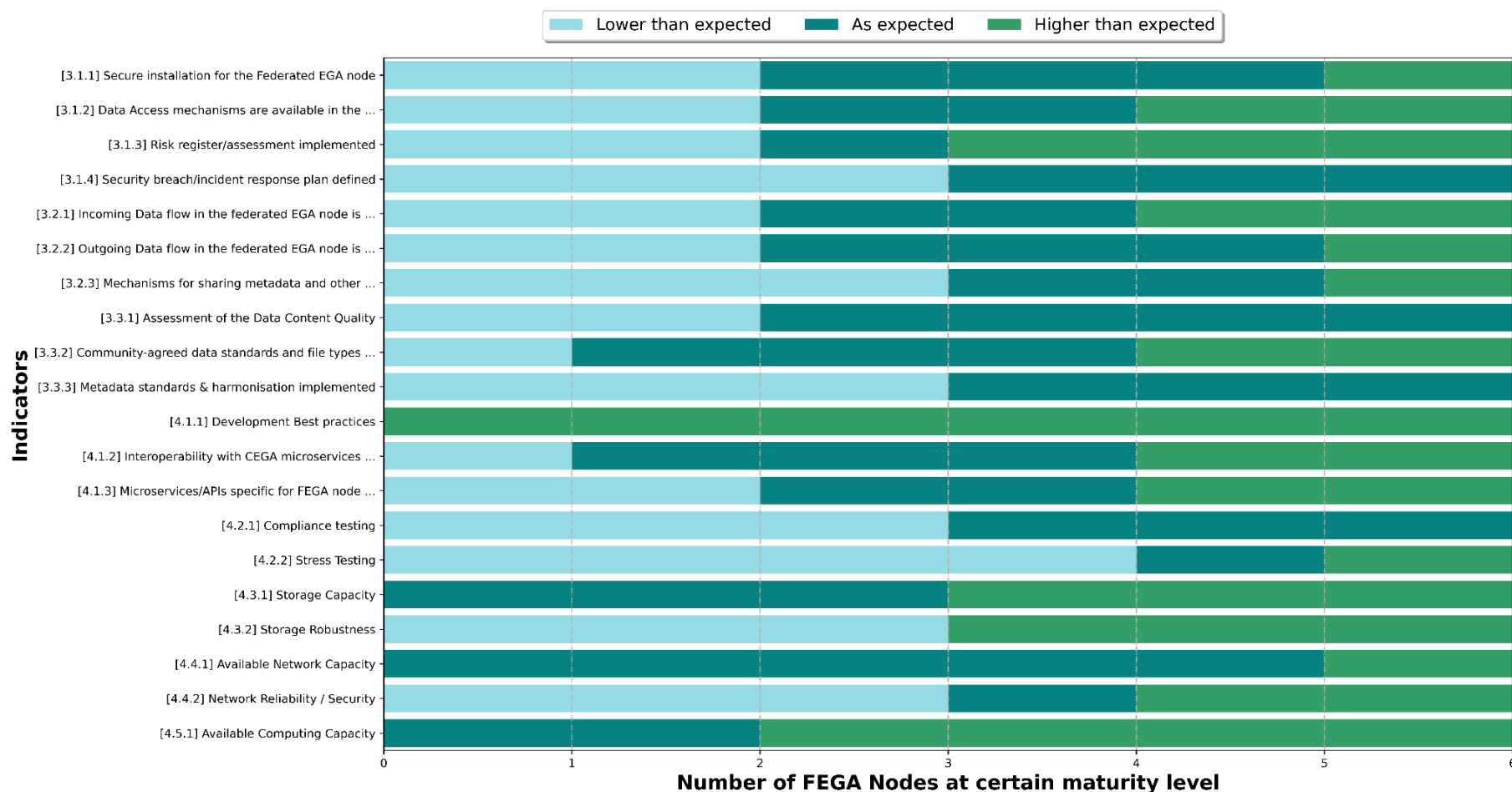
**Figure 3.** Aggregated results for seven Nodes taking part in the initial evaluation of the FEQA, including the expected level for each individual indicator within their corresponding domain/subdomain.





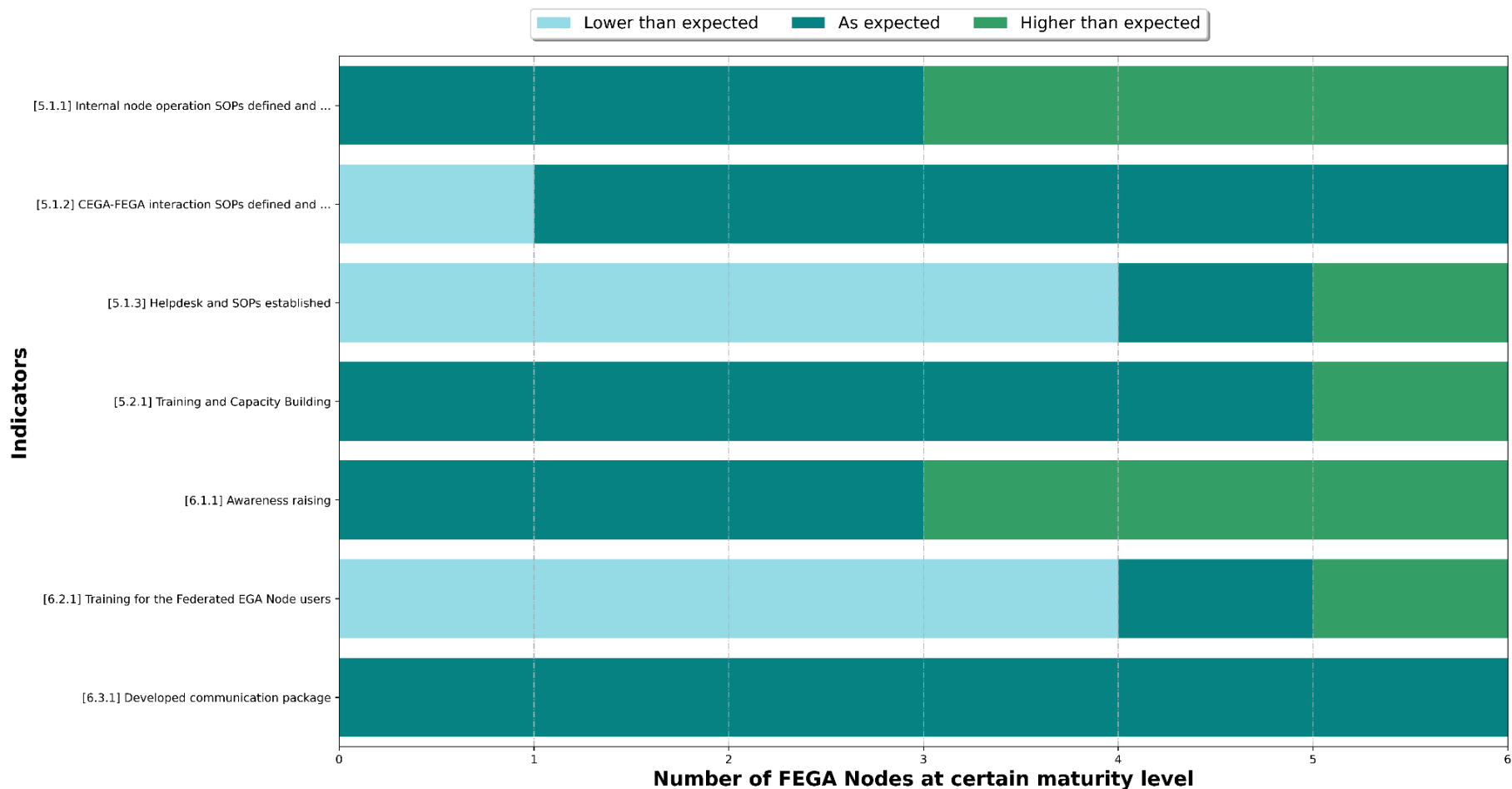
**Figure 4.** Aggregated results for FEGA Nodes, e.g. discarding Central EGA for this dedicated analysis, when considering only the first and second FEGA Maturity Model domains.





**Figure 5.** Aggregated results for FEAGA Nodes, e.g. discarding Central EGA for this dedicated analysis, when considering only the technical, e.g. Data & metadata management, Technical Infrastructure, domains.





**Figure 6.** Aggregated results for FEGA Nodes, e.g. discarding Central EGA for this dedicated analysis, when considering the remaining domains, e.g. Operations Support & Communications, Community Building, and Engagement.



## 5.2 Federated EGA Knowledge Base

[Federated EGA Knowledge Base](#)<sup>10</sup> is a reference point to share documentation, best practices and other relevant documentation for setting up and operating a Federated EGA Node.

The Knowledge Base materials provide guidance for establishing a Node in the Federated EGA Network. The materials are based on the knowledge and experiences of current Nodes and their particular use cases. The new Node's development might differ depending on their use cases and mandates from stakeholders.

Federated EGA Knowledge Base is structured with the same sections as Maturity Model:

- **Governance and Legal**, the information in this section covers topics related - but not limited - to Federated EGA Node governance, strategic planning, sustainability, and legal operating framework. A highlight of this section is the [Federated EGA Collaboration Agreement](#)<sup>11</sup>, the signing of which represents formal inclusion in the FEAGA Network.
- **Data and Metadata management**, the information in this section covers topics related - but not limited - to best practices for data security, data access management, (meta)data standards, and data flow.
- **Technical and Operational**, the information in this section covers topics related - but not limited - to technical infrastructure, testing, software/hardware, SOPs, Helpdesk, and team capacity building.

Highlighted items of this section include:

- [Federated EGA Node Operations Guidelines v2.0](#)<sup>12</sup>
- [Roadmap for Federated EGA Node end-to-end demonstrator](#)<sup>13</sup>
- **Outreach and Training**, the information in this section covers topics related to communications, or training aspects of establishing a Federated EGA Node.

## 5.3 Standard Operating Procedures (SOPs)

Standard Operating Procedures (SOPs) for running a Federated EGA Node are based on a template document: [TEMPLATE Federated EGA SOP - v1.0](#)<sup>14</sup>

The SOPs proposed for the Nodes of the Federated EGA are based on the examples extracted from the different SOPs currently used in the Central EGA for the three main categories:

### 5.3.1 User-facing Processes<sup>15</sup>

- Handling Incomplete Submissions (example)
- Handling Unresponsive DACs (example)
- Users Sharing Credentials (example)
- Reporting Unreasonable Users (example)

<sup>10</sup> <https://ega-archive.github.io/FEAGA-onboarding/>

<sup>11</sup> <https://ega-archive.github.io/FEAGA-onboarding/topics/governance-legal/#fega-collaboration-agreement>

<sup>12</sup> <https://ega-archive.org/files/EGA-Node-Operations-v2.pdf>

<sup>13</sup> <https://ega-archive.github.io/FEAGA-onboarding/topics/technical-operational/#3-evaluate-your-implementation>

<sup>14</sup> <https://ega-archive.github.io/FEAGA-onboarding/topics/technical-operational/#standard-operating-procedures-sops>

<sup>15</sup> <https://ega-archive.github.io/FEAGA-onboarding/topics/technical-operational/#example-user-facing-process-sops>

- Changing User Institution (example)

### 5.3.2 Internal Node Processes<sup>16</sup>

- EGA Helpdesk Ticket Management (example)
- EGA Helpdesk Ticket Assignment (example)

### 5.3.3 Joint CEGA/FEGA Node Processes<sup>17</sup>

- Federated EGA Metadata Check and Release Protocol
- Federated EGA Request from submitter to submit to FEGA Node

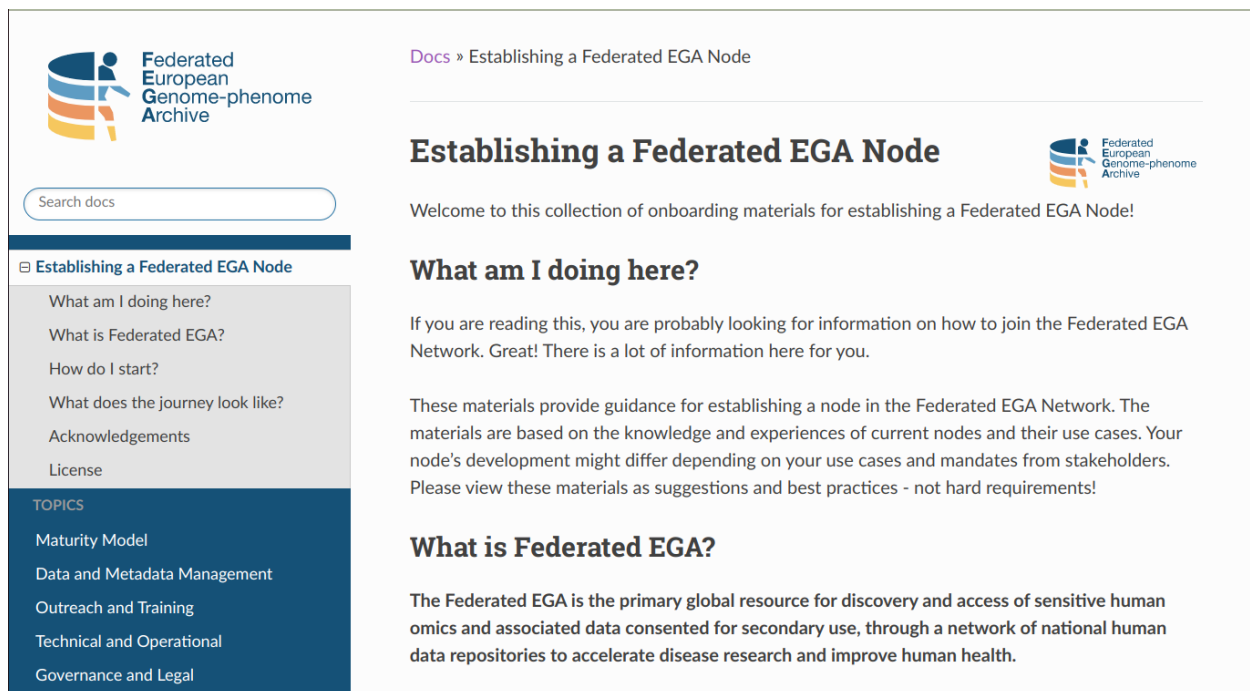


Figure 7. Screenshot of the Federated EGA Knowledge Base.

## 6. Conclusions

ELIXIR Converge WP7 activities have been centred in accelerating the formal creation and deployment of the Federated EGA Network of Nodes. In order to achieve this ambitious goal, it has been necessary to develop documentation and guidelines supporting the deployment of Federated EGA Nodes beyond the already engaged ones. Indeed, the materials generated and provided as part of this deliverable - especially the FEGA Maturity Model - can be taken as a roadmap for the establishment, deployment and maintenance of operational Nodes of the Federated EGA.

The FEGA Maturity Model represents an important asset built after parallel experiences across other projects in ELIXIR, e.g. 1+MG/B1MG and FAIRplus. Reflecting on different aspects helps to set up priorities when establishing a Node and also contributes to periodically reviewing existing indicators.

<sup>16</sup> <https://ega-archive.github.io/FEGA-onboarding/topics/technical-operational/#example-internal-Node-process-sops>

<sup>17</sup> <https://ega-archive.github.io/FEGA-onboarding/topics/technical-operational/#shared-fega--cega-sops>

Both technology and regulation tend to evolve over time, so every FEAGA Node evolves as well. The initial work done with the five inaugural Federated EGA Nodes, the Central Node and a seventh Node willing to join the FEAGA is quite significant, as it offers real experiences on the usability of the Maturity Model as a self-guiding tool and puts the focus on areas for improvement either at the FEAGA level or at the Nodes level.

In the same way, a self-assessment tool has been created to evaluate the different indicators of the Maturity Model. This tool becomes an important resource by allowing a continuous improvement of the Nodes, as they can validate their responsibilities, coordinate their status with the federation committees and find improvement points for their implementation. Results show the differences between established Nodes and new Nodes willing to join FEAGA. It also opens the door to future iterations and refinements on the expected level based on the Federated EGA Node maturity. Different expectation levels might be set for more mature Nodes. Indeed, this has already been observed for Central EGA where most of the indicators have the highest possible rank (level 5).

The Federated EGA Knowledge Base is the perfect complement to the FEAGA Maturity Model as it represents the place where documentation and training materials are shared across the FEAGA Nodes. Documentation includes explanation of the FEAGA Maturity Model as well as reference SOPs for the operation of FEAGA Nodes based on existing experiences at Central EGA. It is expected that this knowledge base will evolve over time as further materials are added, more FEAGA Nodes join the network, and existing assets, e.g. FEAGA Maturity Model, are re-evaluated and extended.

After a year and a half of work, ELIXIR Converge WP7 has created a productive team which includes different representatives of national Nodes in production and personnel specialised in creating maturity models. The working method used to create the Federated EGA Maturity Model has been exhaustive, and the participation of the different members of the team in the diverse internal activities has been very successful. This has made it possible to form a group of people capable of transferring the knowledge acquired to collaborate with other WPs in their maturity model development processes, e.g. WP9.

## 7. Impact

Activities related to this deliverable have a direct impact on the growth of the Federated EGA Network. In the specific case of the Federated EGA Maturity Model and its self-assessment tool, it will be used as a validator by the different committees of the Network when incorporating new Nodes. Thus, when a national Node applies to connect to the Network, it must be mature enough to start the different procedures.

A greater impact of the effort done to accomplish this deliverable is expected in the European Genomics Data Infrastructure (GDI) as Federated EGA will be a core component of this project. Indeed the Federated EGA Maturity Model and the Knowledge Base, including the different SOPs and templates, are instrumental in meeting the deployment objectives of the GDI (>15 operational nodes in 2026).

On the other hand, after the first evaluation of seven Nodes (five inaugural FEAGA Nodes, a future Node and the Central EGA itself), some aggregate results have been obtained that have made it



possible to identify improvements in the model itself and show in which domains there are more strengths and weaknesses in the Federated EGA Network as a whole.

In addition, after presenting the work carried out to complete this deliverable in different meetings, the visibility of the Federated EGA Network has increased, creating the interest of new possible Nodes (even outside Europe, e.g. Canada or Australia) as well as of other WPs in CONVERGE regarding developing a similar process to incorporate Nodes - with data management and computational capabilities - for viral sequences handling.

## 8. Next Steps

The perspective of the work is:

1. Improve the specific aspects of the associated documentation to the Federated EGA Maturity Model. This effort will focus on adding extended descriptions, examples and clarifications of those indicators that, in its first version, have generated more confusion.
2. Continue improving the documentation available at the Federated EGA Knowledge Base, with a focus on particular experiences at individual FEAGA Nodes.
3. Work collaboratively to Increase the number of available SOP templates.
4. Start the conversation on the need to have different expectation levels for a FEAGA Node depending whether it is a well established one or a new one joining the network.
5. Propose a disengagement model for those cases where a given FEAGA Node cannot continue being part of the network.

## 9. Deviation from Description of Action

No major deviation from the expected Description of Action has occurred for this deliverable as the creation, evaluation and use of the Federated EGA Maturity Model is a necessary step for the establishment of a network of FEAGA Nodes capable of managing COVID-19 host genomics data that is exposed through the COVID-19 Data Portal.