

Collaboration Activities between EGI and WLCG

DRAFT v0.99a, 19 April 2021

Editors: M. Litmaath (WLCG), M. Viljoen (EGI)

[1. Introduction](#)

[2. Coordination and Communication](#)

[3. Support](#)

[4. Services and capabilities](#)

[5. Forums, Events and Dissemination](#)

[6. Service evolution](#)

[7. EOSC and HPC](#)

[8. Environmental impact](#)

1. Introduction

The purpose of this document is to outline the areas where the EGI Federation and WLCG are actively collaborating and have been doing so over a period of 10 years, as well as common areas where new opportunities may arise for further collaboration in the changing IT landscape.

The **EGI Federation** is an international infrastructure that provides advanced computing and data analytics services for research and innovation. The federation is composed of providers from national/community initiatives forming one of the largest distributed computing infrastructures for researchers in the world. The federation delivers technical and support services at European and global scale, including high-throughput and cloud computing, storage and data management, analytics, consultancy and support, training and co-development.

The EGI Federation is coordinated by the **EGI Foundation**, an organisation with headquarters in Amsterdam, the Netherlands, whose objective is to coordinate and develop, in collaboration with its participants, the EGI infrastructure that provides long-term distributed compute and storage resources for performing research and innovation activities.

The foundation offers a federation and management platform that enables service providers to harmonise interfaces and connect to a common hub, while engaging with research communities to understand the demand, simplifying the access and driving innovation together. At the top level, the EGI Federation is governed by the participants represented in the EGI Council. An Executive Board is charged with the management of the foundation in collaboration with the chair of the EGI Council and with the Director of the EGI Foundation. Further details are available from www.egi.eu.

The **Worldwide LHC Computing Grid** (hereafter referred to as WLCG) is a global collaboration that links grid infrastructures and computer centers worldwide. Its main goals are the storage and analysis of the immense amounts of data generated by the four main experiments at the Large Hadron Collider (LHC) at CERN. WLCG is a mature research community, both in its own right and in its use of grid technology. Working closely with EGEE and other projects, WLCG pioneered the (co-)development of a number of scientific computing and research strategies, policies and tools that remain in use today. In recent years of LHC operation, the experiments have produced more than 100 PB of raw, derived and simulated data annually, which many thousands of scientists around the world access and analyse. The total amount of data accumulated over the past decade exceeds 1 EB and the amount of CPU cores routinely accessible to WLCG approaches 1M. More than 160 computer centers around the world formally pledge resources to WLCG, while O(100) sites contribute through other arrangements. About two thirds of the total capacity is provided by sites affiliated with EGI. Their contributions, also used by other VOs, are as of early 2021 estimated to make up at least three quarters of the computing and storage resources in EGI. To be entitled to the reserved status of WLCG Tier-1 or Tier-2, sites need to be able to make use of various EGI services, in particular for accounting. Such sites either have to be members of EGI or be recognized through formal partnerships between EGI and other infrastructures. Further details at wlcg.web.cern.ch.

2. Coordination and Communication

The EGI Federation and WLCG may coordinate common activities through participation in the EGI [Operations Management Board](#) - an advisory body which develops technical and strategic priorities for the EGI federation - and the WLCG [Management Board](#) - the executive body overseeing the current operation of the WLCG infrastructure, while technical matters are usually handled by operations experts on either side.

EGI and WLCG have several communication channels for keeping each other informed on relevant activities and for discovery of further areas that offer the potential for cooperation. These channels are detailed in a subsequent chapter. Both parties have contact persons in their respective operations coordination teams who oversee current and future cooperative activities.

3. Support

WLCG users and site managers are supported through a range of crucial services detailed in the next chapter. In particular, the EGI Federation is in charge of managing the [GGUS](#) helpdesk system support units, ticket triage, their escalation and first- and second-level support. The EGI Federation has been providing funding through projects and fees to maintain, evolve and upgrade GGUS to introduce novel functionality in the helpdesk. GGUS is operated by KIT according to terms and conditions defined in an Operational Level Agreement.

Many infrastructure services are built from products originating in WLCG institutes and while their initial purpose or scope may typically have been limited to WLCG, they subsequently were adopted by many other communities. Most such products are various types of middleware that are crucial for large-scale grid operations. Examples include [VOMS](#), [INDIGO-IAM](#), CREAM (discontinued), [ARC](#), [DIRAC](#), [CVMFS](#), [dCache](#), [DPM](#) (deprecated), [StoRM](#), [XRootD](#), [FTS](#), [Rucio](#) and their respective client tools. The maintenance of the Globus middleware was taken over by the [Grid Community Forum](#) (Grid CF), a collaboration of several institutes whose supported communities currently depend on Globus in some ways. Most of those institutes are particularly concerned with the remaining use cases in WLCG, but their efforts have benefited many other user communities in EGI as well. The Grid CF initiative is therefore supported by both WLCG and EGI. Further examples of collaboration are the [HEP-SPEC06](#) benchmark, to be superseded by the modern [HEP-Score](#) suite under construction, and even [Indico](#).

The institutes and collaborations developing and maintaining such products have also supported other communities throughout the years, often through funding from EU projects, and generally will continue to do so, constrained by the level of resources that can be made available for such purposes. If for a given product a guaranteed level of support better than best-effort becomes essential, a specific arrangement would have to be made with the institute or collaboration behind that product. The wider adoption of products originating from WLCG institutes constitutes spin-off from investments in HEP. The EGI Federation facilitates such knowledge transfer by setting up example instances as catch-all services for other communities and by organizing webinars and other training events concerning particular products of potential interest to user communities or resource center administrators.

4. Services and capabilities

The EGI Federation and WLCG each have their own portfolios of services and capabilities. Here we list those that have overlapping interests and benefits. The following services are developed in coordination with peer infrastructures in the world with the aim of ensuring interoperation. Such infrastructures include [Open Science Grid](#) in the USA, [Compute Canada](#), [NICIS](#) in South Africa, and [ASGC](#) in Taiwan who coordinate the Asia Pacific component of the EGI Federation.

4.1 Infrastructure and operations services

- **Infrastructure catalogue** a.k.a. Configuration Database ([GOCDB](#)). This includes information on services which is suitable for both human consumption and automatic consumption via APIs. Available information includes service endpoints, scheduled downtime as well as staff involved in running services (system administrators and security contacts).
- **Accounting Repository and Portal** for CPU ([APEL](#)). EGI Accounting stores user accounting records from various services offered by EGI, such as Cloud, HTC and storage usage. It works thanks to message brokers that transfer usage data from the services to a central repository of information, which then can be visualized in the [Accounting Portal](#). Note that WLCG *storage* accounting is being handled through WLCG services.
- **Helpdesk Service** ([GGUS](#)). Central frontend to a distributed helpdesk infrastructure which interfaces to ticket systems of sites and Technology Providers. From 2021, work will begin on developing a new domain-agnostic helpdesk solution as part of [EOSC Portal](#) work. When it is sufficiently mature, the current EGI Helpdesk is expected to be migrated to this new solution once it has been demonstrated to meet all EGI Helpdesk requirements, including those from WLCG.
- **Helpdesk Support Staff** for first- and second-level regular ticket triage and processing, with assignment to third-level technology provider experts, sites or service managers as needed. First- and second-level support are funded through EGI.
- **Availability Monitoring** ([Nagios](#) and [ARGO](#)). This service supports monitoring and access to historical monitoring data for the infrastructure. It is complementary to WLCG monitoring. Its results are overseen by EGI Operations and unresolved issues are followed up through the Helpdesk second-level support as needed. Monitoring relies on a distributed messaging infrastructure for the exchange of monitoring probe results.
- **Operations Portal** bundling different functionality including dashboards used by EGI Operations to help ensure that the infrastructure is in good shape. The [Operations Portal](#) also includes VO management facilities, a security dashboard and a broadcast tool to send information targeted at different communities.
- **EGI AAI Proxy** ([Check-in](#)). Check-in is a fully managed AAI platform enabling integration of different identity providers, service providers, attribute authorities, community AAI and infrastructure proxies, thereby facilitating access to several other services mentioned in this chapter. Its importance to WLCG will increase as users and resource center administrators gradually move away from X509 user certificates toward federated identities.

4.2 Software distribution services

- **UMD/CMD Quality Assurance Services** are the EGI software distribution channels. The [Unified Middleware Distribution](#) provides additional assurance and verification over other software distribution channels e.g. via its Staged Rollout procedure.
- **Deployment and Operations Documentation** ([EGI Collaboration Space](#), [EGI Documentation](#), [Document Database](#) and [Wiki](#)) to assist with infrastructure integration and service documentation.
- **Middleware deployment campaigns** are complementary to WLCG efforts and followed up by EGI Operations and second-level support. These activities help with upgrade and migration campaigns which are followed up directly with affected sites using the Helpdesk and, where necessary, tracking problems with product teams.

4.3 Software support

Many of the services within EGI originate within High Energy Physics institutes and are being actively developed by the WLCG community with non-HEP use cases being supported by the developers. This includes the services mentioned earlier in this document.

The EGI Foundation contributes to the maintenance, support and/or improvement of a subset of these software projects. Examples include(d): [BDII](#), CREAM (discontinued), [CVMFS](#), [DIRAC](#), [DODAS](#), [DPM](#) (deprecated), [FTS](#) and [INDIGO-IAM](#).

4.4 Security services and activities

The establishment and retention of trust are fundamental to any distributed computing infrastructure. The EGI Federation and WLCG are working closely together on matters pertaining to security. Common activities and services include the continual evaluation of security threats, operational security and security incident response and, particularly through hands-on training at EGI conferences and other events, the dissemination of security best practices. The EGI Computer Security Incident Response Team ([CSIRT](#)) is jointly funded by both EGI Foundation and WLCG and is crucial to the operations of WLCG and EGI. It includes experts in the field of IT Security who support EGI and WLCG sites to deal with ongoing security attacks, offer global security incident response coordination, provide in-depth expertise to defend the infrastructure, evaluate new threats that become known, monitor for vulnerabilities in the e-Infrastructure via dedicated security monitoring and provide an appropriate response to each case under consideration. Any potentially relevant vulnerabilities can be reported by anyone to the EGI Software Vulnerability Group ([SVG](#)), who will analyze each case and advise the CSIRT, potentially affected resource and/or technology providers and/or user communities on an appropriate course of action. The SVG is mainly concerned with specific software that is used to enable distributed computing for EGI communities, in particular grid middleware. The SVG largely relies on expert volunteers from EGI and WLCG communities. The follow-up of sufficiently serious vulnerabilities includes notification of affected sites, monitoring the upgrading/patching of affected software and ultimately acting to suspend sites who fail to take proper action. The CSIRT also has regular contact with security experts in other fields, which helps ensure EGI and WLCG are at the forefront in dealing with security threats. The team participates in the international [WISE Community](#), representing EGI and WLCG interests.

The team endeavours to run a comprehensive security challenge at least once per year, to test the readiness of the community to deal with security incidents in scenarios that are designed to be as realistic as possible. The analyses of these security challenges are used to keep improving the effectiveness of dealing with any types of security threats.

EGI represents [EUgridPMA](#) in the Interoperable Global Trust Federation ([IGTF](#)) and is in charge of the periodic release of the EGI Trust Anchors based on the updated IGTF Accredited CA distribution. To ensure interoperability within and outside of EGI, the Policy on Acceptable Authentication Assurance defines a common set of trust anchors that all sites in the EGI Federation should install.

Finally, security experts from both EGI and WLCG are collaborating to help ensure the harmonization across multiple domains of security policies and technologies, to the benefit of users and resource providers, for example through activities under the WISE Community umbrella.

5. Forums, Events and Dissemination

Both parties participate in each other's events such as workshops and conferences where such is beneficial in order to further the activities outlined in this document. Participation implies either physical attendance or monitoring meeting minutes/communications etc. for relevant topics and following up as needed. EGI events and boards include:

- The [Operations Management Board](#) (OMB) meetings and mailing list serve to facilitate maintenance of operations, policies, NGI communication and security. Furthermore, the OMB also allows concerns about key products and corresponding requirements from user communities to be brought up for discussion and follow-up.
- The [Security Policy Group](#) (SPG) provides policies that define the expected behaviour of sites and users for a secure distributed computing infrastructure.
- The [Technology Coordination Board](#) (TCB) coordinates the identification, testing, validation, adoption, provisioning, and decommissioning of existing and future technologies relevant to the delivery of the EGI Service Portfolio, including operations tools of common interest to the EGI Federation and WLCG.
- The [User Community Board](#) (UCB) provides a forum for end-users in structured user communities to shape the infrastructure for their needs.
- [Operations meetings](#) serve to record the overall status of the infrastructure, discuss ongoing issues and present plans for short-term changes, e.g. concerning middleware.
- UMD Release Team ([URT](#)) meetings serve to discuss middleware issues, track the verification of new versions and plan their releases in the EGI Software Repository.
- EGI [conferences and workshops](#).

WLCG forums and boards include:

- [Grid Deployment Board](#) (GDB) meetings mainly serve to present and discuss options and plans for the technical evolution of the WLCG infrastructure, taking stock of developments in industry, academia, partner organizations, national and international projects.
- The [Management Board](#) (MB) is an executive body overseeing the current operation of the WLCG infrastructure and taking high-level decisions on its evolution, in particular in matters of policy.
- Weekly [Operations meetings](#) serve to record and potentially discuss short-term matters.
- Monthly [Operations Coordination](#) meetings are concerned with longer timescales, particularly involving task forces and working groups.
- The WLCG Data Organization, Management and Access ([DOMA](#)) project is an R&D project to construct and implement a roadmap for the data management aspects of the infrastructure toward readiness for the requirements of the High-Luminosity phase of the LHC, which is to start in 2027.
- The WLCG [Authorization Working Group](#) is tasked with the evolution of authentication and authorization paradigms used by WLCG toward modern technologies used in industry as well as academia.

Dissemination serves to increase the visibility of activities of EGI and WLCG, in particular concerning the areas of collaboration between both parties. Dissemination channels include events, newsletters, websites and reports. For example, existent collaboration synergies were presented and discussed in a Grid Deployment Board meeting on the WLCG side and in an Operations Management Board meeting on the EGI side, both in September 2020.

In matters of dissemination it is important to make clear how either party depends on the other and how there are mutual benefits. WLCG is the largest user community in terms of scale of its data processing and analysis needs in the EGI Federation. WLCG sites still provide the vast majority of the resources making up the EGI infrastructure. WLCG institutes also provide a large range of middleware and other products to the benefit of many other user communities and sites in EGI. On the other hand, WLCG has delegated to EGI the responsibility for a range of crucial services and activities that are not specific to WLCG, but also serve many other participants in EGI. WLCG thus depends on continuity in the vital areas managed by EGI.

6. Service evolution

Both EGI and WLCG look to phase out legacy protocols and services, while introducing new technologies commonly employed in industry and/or academia. In the area of AAI (Authentication and Authorization Infrastructure) the aim is that end users will be able to access distributed computing services directly using just their home institute credentials, obviating the need for cumbersome X509 user certificates. Through the [AARC](#) projects, EGI and WLCG have established common ground for that paradigm shift. During a transition period that probably will last through the whole of Run 3, the [Grid Community Forum](#) collaboration is set to maintain Globus legacy middleware that remains critical for the use of X509 certificates in EGI, WLCG and partner communities. EGI is a major contributor to the Grid CF. Specific data or job management activities will gradually start making use of tokens based on industry standards instead of X509. The WLCG [DOMA](#) R&D project is concerned with guiding major storage and data management systems toward supporting these new AAI paradigms as well as HTTPS / WebDAV for third-party data transfers. On the computing side a strong increase is expected in the use of container images distributed through [CVMFS](#). Containers will also be increasingly used as a basis for site services, orchestrated e.g. through Kubernetes. The configuration and operation of site services are an area where simplification remains desirable, for example through the sharing of Ansible playbooks in a common repository managed by EGI Operations.

EGI Foundation will continue looking to introduce other user communities to spin-off technologies from WLCG experiments and institutes. Important examples are [CVMFS](#), [DIRAC](#), [FTS](#) and [Rucio](#), where EGI also provides funding for maintenance. While in general the technical support to new communities will be best-effort, EGI can provide funding for development and support in specific cases. For example, EGI Foundation is allocating major funding for DIRAC development and support of communities beyond HEP.

Another example concerns the support of the [DPM](#) storage middleware, which not only has been serving WLCG for many years, but also has many instances for other user communities in EGI. As the effort available for development and support of the DPM has decreased in recent years, EGI Foundation has stepped forward and set up an arrangement to fund additional expertise to facilitate the migration of DPM instances to alternative solutions. EGI Foundation has also provided support to sites in their migration from discontinued CREAM CE services to other solutions, in particular through the provision of training events and workshops.

7. EOSC and HPC

EGI Foundation has been and is very much involved in the development of EOSC and the [EOSC Portal](#) via the completed [EOSC-hub](#) project and the [EOSC-Future](#) project, which started in April 2021. EGI Foundation is also being funded to further develop the distributed and cloud computing service offering of resources to the EOSC Portal via the [EGI-ACE](#) project, which will develop the EOSC Compute Platform based on a distributed hybrid infrastructure incorporating Cloud IaaS and HPC as well as HTC.

CERN and other WLCG institutes have been partners in a number of EGI Foundation flagship projects. Within the EGI-ACE project, CERN participates in the HPC Integration task, where WLCG requirements are to be incorporated as part of use case analysis and WLCG needs should be met in developing interoperability guidelines for HPC with the EOSC Compute Platform. EGI and WLCG may have further opportunities to work together in the steadily evolving HPC landscape where new projects continue to emerge. Security is one area where a strong collaboration between EGI Foundation and HPC centers has already been very fruitful.

8. Environmental impact

Limiting environmental impact and carbon footprint via green computing (GC) is a relatively new and increasingly important concern relevant to many areas of the EGI Federation and WLCG, with various prospects. It relies significantly on expertise and the successful application of best practice. EGI Foundation and WLCG can expect opportunities in nurturing and sharing best practice concerning the following aspects, particularly through the [EGI-ACE](#) project:

- data center management (e.g. PUE computation and improvement);
- software (e.g. optimised algorithms for minimising resource utilisation);
- GC aware procurement of hardware etc;
- efficient networking and data transfer paradigms;
- energy conscious management involved in service delivery;
- responsible energy procurement (electricity, gas, renewable);
- data center/machine room building design and construction;
- social aspects and promotion of green thinking.