GOCDB - new communities, new requirements, new architecture

Greg Corbett[1,*], *George* Ryall[1] and *Ian* Collier[1]

[1]Science and Technology Facilities Council, Scientific Computing, Didcot, United Kingdom

Abstract. GOCDB is the official repository for storing and presenting EGI and WLCG topology and resource information. It is a definitive information source, with the emphasis on user communities to maintain their own data. It is intentionally designed to have no dependencies on other operational tools for information.

In recent years, funding sources and user communities have evolved and GOCDB is developing to meet the resulting new requirements, such as allowing more programmatic updates to the data within GOCDB. We will explain the roadmap for developing GOCDB and the motivation for the changes on that Roadmap.

As GOCDB, and the team supporting it, has evolved, we have re-examined the underpinning architecture for GOCDB. We will set out the changes we have made as a result.

1 Introduction

The Science and Technology Facilities Council (STFC), part of UK Research and Innovation (UKRI), provides the Grid Operations Centre Database (GOCDB) for EGI [1], co-funded by EGI.eu [2] and EOSC-hub [3]. The GOCDB is the official repository for storing and presenting EGI and WLCG [4] infrastructure topology and resource information.

The GOCDB is a definitive information source where data is directly populated and managed in the system, it provides a web portal for users to edit their own data and read other data and an API for operational tools to programmatically access the data. GOCDB has been a production service since 2004 and over the course of several EU projects (EGEE phases 2 [5] and 3 [6], EGI-Inspire [7], EGI-Engage [8] and EOSC-hub) funding sources, user communities and the team supporting the service have evolved and GOCDB is developing to meet the resulting new requirements. This paper sets out the roadmap for developing GOCDB, the motivation for the changes on that roadmap and the changes we have made as a result.

2 Grid Operations Centre Database

The Grid operations Centre Database (GOCDB) is the official repository for storing and presenting EGI and WLCG infrastructure topology and resource information.

It is a definitive information source where data is directly populated and managed in the system. Because GOCDB is a primary data-input source, the portal applies a range of business rules and data-validations to control input. It applies a comprehensive Role-based authorization model that enables different actions over different target resources. The Role model allows communities to manage their own resources where users with existing roles can approve or reject new role-requests.

It is intentionally designed to have no dependencies on other operational tools (other than the EGI Check-in service). For example, it does not query other systems to populate its core data model. The underling Oracle database is hosted by the STFC Database Services Team with nightly tape backups. An additional failover instance is hosted at a second STFC site (Daresbury Laboratory). The failover instance is synchronized hourly against the production data.

---

[*] Corresponding author: greg.corbett@stfc.ac.uk

Entities in GOCDB:
- are arranged into a hierarchical tree, shown in Fig. 1, of national federations of shared computing resources (NGIs) [9], sites, services and service endpoints
- contain information such as contact information, users with roles over the entity and a list of downtimes
- can be "scoped" to allow for more flexible categories and groupings, such EOSC-hub, EGI and WLCG.

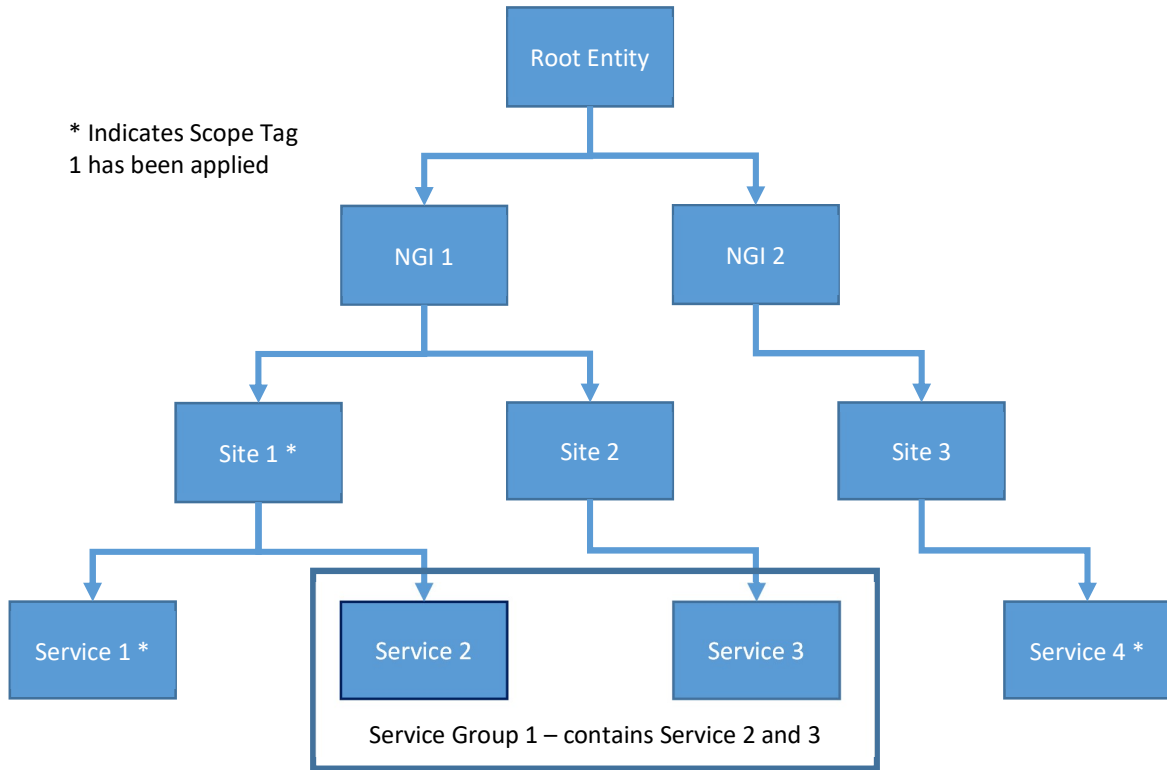GOCDB provides a web portal for users to edit their own data and read other data.

Fig. 1. The hierarchical tree of entities within GOCDB.

2.1 GOCDB PI

GOCDB also provides an API for operational tools to programmatically access the data, which is well documented [10, 11]. API calls can be either Public, Protected or Private:
- public methods (no critical information, no personal mail/details)
- protected methods (contain mails or personal details) – needs an X509 certificate from trusted CA
- private methods (contains security or critical information) – certificate DN needs to be register in GOCDB for authentication

2.2 New Communities

The European Open Science Cloud (EOSC) is Europe's vision of an Open Science as a driver for enabling a new paradigm of transparent, data-driven science as well as accelerating innovation. The EOSC-hub project is one part of the EOSC programme and mobilises providers from the EGI Federation, EUDAT CDI, INDIGO-DataCloud and other major European research infrastructures to deliver a common catalogue of research data, services and software for research [12].

2.3 New Requirements

In recent years, GOCDB's user community has been changing the way they interact with GOCDB, using it to store more dynamic and programmatically updatable data than it was originally designed for. Additionally, the EOSC-hub project has brought new users and requirements to GOCDB.

2.3.1 Automatically creating new Service Types

Services registered with GOCDB must be instances of one of a predefined list of Service Types. These Service Types allow other operational tools to group services together, for example – the APEL accounting tool [13] queries GOCDB for all services of the type "gLite-APEL" to authorise those services to publish accounting data. If a community needs a new service type created, they can requested its creation, this request is then moderated by the wider GOCDB user community before being created by the GOCDB administrators.

Within the EOSC-hub project, there is a need to synchronise Service Types across the projects component infrastructures and their service registries – GOCDB, EUDAT's Data Project Management Tool (DPMT) and EOSC-hub's Service Portfolio Management Tool (SPMT). This will be done by the SPMT publishing a list of EOSC-hub Service Types via an API that GOCDB and DPMT can automatically consume to create Service Types.

2.3.2 Scope Specific View

Within GOCDB, Sites and Services can be "scoped" into flexible categories and groupings. The purpose of these scopes is to allow for segregation of resources within GOCDB without duplicating information. For example, if a site in GOCDB provides different services to EGI and WLCG, this is expressed by applying the "EGI" or "WLCG" (or both) scope to those services, rather than duplicating the site to achieve this split. These scopes can then be specified in API calls to limit the information retrieved – for example, calls made by the EGI monitoring system need only retrieve information with the "EGI" scope. Whilst certain views in the portal can be filtered based on scope, as GOCDB, EGI and EOSC on-board new communities, there is a growing need for GOCDB to offer scope based filtering by default.

As part of the EOSC-hub project, a separate read/write instance of the GOCDB portal and API will be set up, that accesses the same underlying database as the current portal. This view will offer a simplified UI, shown in Fig. 2, with limited customisability, such as: bespoke colour schemes, community specific terminology as shown in Fig. 3, and most views limited to the scope of the portal by default as shown in Fig. 4. This work will allow new communities to benefit from seemingly having their own GOCDB "instances", without actually setting up their own instances and losing the benefits of the current single source of truth. Whilst this will not enforce changes to how current users interact with GOCDB, a key motivator for this work is that resource providers and user only have to Sites/Users only have to interact with one service registry, specific portals could be created for existing GOCDB communities if required, such as WLCG or ELIXIR [14].

Fig. 2. The traditional GOCDB menu (left) and the simplified menu provided by the EOSC-hub scope specific view.

Fig. 3. The traditional terminology ("Service Endpoints", left) and the EOSC-hub terminology ("Service Components", right)
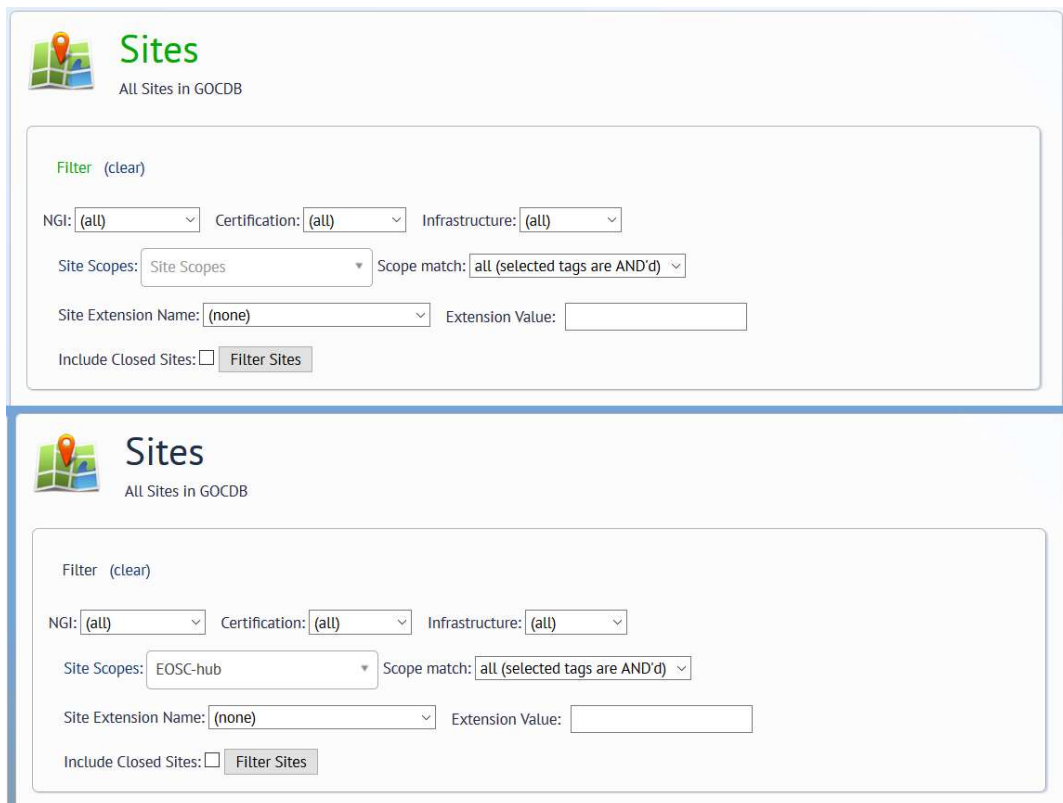
Fig. 4. The traditional, unfiltered view of sites in GOCDB (top) and the pre-filtered view of sites provided by the EOSC-hub scope specific view.

2.3.3 Extensions to the Write-API

When the service was first designed, its use case was to store mostly static data with infrequent updates. In recent years, this has begun to change as communities seek to make programmatic updates to the data within GOCDB. This lead to the development of a Write-API to allow programmatic CRUD operations on community defined extension properties. This Write-API will be extended to allow for the creation of services automatically and CRUD operations on service endpoints.

2.4 New Architecture

The historical architecture of the GOCBD service was a single legacy virtual machine directly accessible to the outside world via an egi.eu DNS entry, as shown in Fig. 5, along with an offsite, read only, failover. As the team supporting GOCDB has evolved the underpinning architecture of GOCDB has been re-examined. As a result of this, the decision was made to move the GOCDB service "closer" to the RAL WLCG Tier 1 [15].

The first steps of this migration, as shown in Fig. 5, have already been undertaken. The legacy production host has been moved behind two highly available proxies provided by the Tier 1. This move made the GOCDB service IPv6 accessible for the first time, a key requirement from WLCG, without changes to the legacy production host. The proxies' configuration has also increased the service reliability and availability by allowing the service to automatically fall back to the offsite failover in the event of disruptions to the legacy production host. Previously, the failover would have only activated for serious disruptions to the service, on the order of 2 days of downtime, due to the potential disruption caused by changing egi.eu DNS entry.
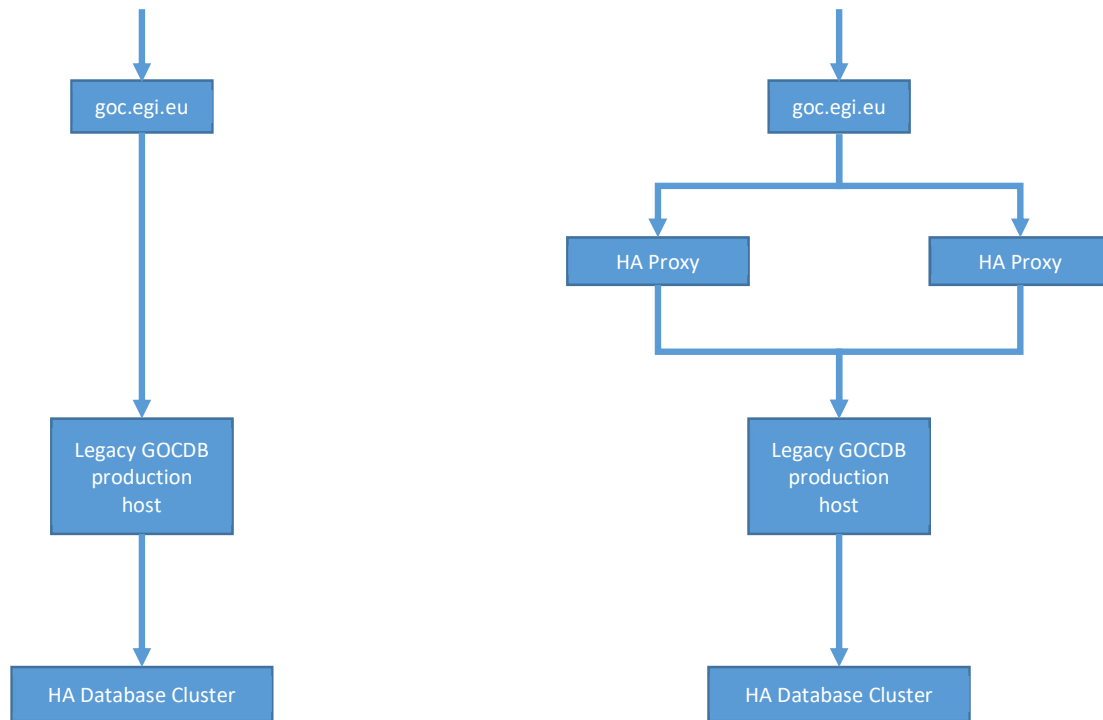
Fig. 5 The historical GOCDB service architecture (left) and the current GOCDB service architecture (right).

In future, GOCDB will itself be moved to a highly available setup, having two configuration managed hosts providing the service and the current production machine will be decommissioned, this setup is shown in Fig. 6.

This change in architecture will bring benefits to both the "service team delivering GOCDB and the users of GOCDB. The move to the Tier 1 infrastructure provides an opportunity to change the underlying operating system of GOCDB. Currently GOCDB is running on Red Hat 6, which enters "Maintenance Support 2 Phase" in November 2020 [16]. As well as this, the recent Red Hat licensing changes make it undesirable to migrate the service to Red Hat 7. As such the service will move to Scientific Linux 7. The service can also take advantage of the Tier 1's configuration management system, making the service easier to manage by ensuring that the multiple hosts that will provide it are configured identically. Configuration management will make future operating system upgrades easier and also allows for a pre-production environment to be set up[†] using the same templates as used in production.

Making GOCDB highly availably will allow many updates to GOCDB to be transparent to the user, increasing the service's availability and reliability.
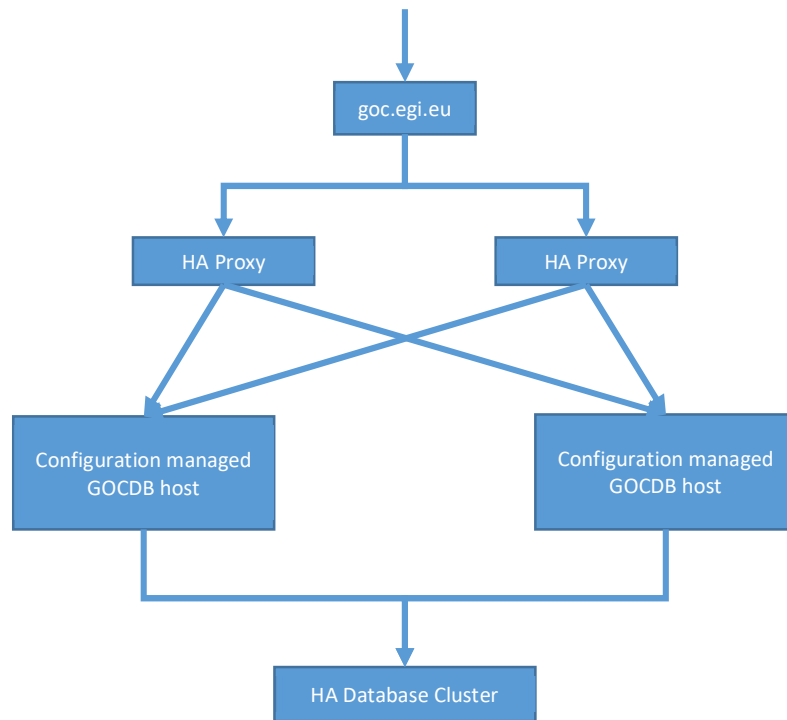

[†] https://gocdb-preprod.egi.eu/portal/

Fig. 6. The future GOCDB service architecture

3 Beyond the Current Roadmap

The EOSC will bring an increase of between a factor of two to an order of magnitude of service endpoints, data stored and transactions. To better serve this increase of requests efficiently, we will seek to replace our "homebrew" Model, View, Controller with a modern and efficient framework.

4 Summary

Over the course of several projects, the funding, requirements and the team that deliver GOCDB has evolved and will continue to do so. During the current project, EOSC-hub, GOCDB will move to a more resilient architecture, expand its Write-API to allow more programmatic access to its data, providing scope specific views of its data, and better support future EOSC use cases by fetching new Service Types from the EOSC-hub Service Portfolio Management Tool API.

References

1. EGI.eu, "EGI | About EGI," [Online]. Available: https://www.egi.eu/about/. [Accessed 04 March 2020].
2. EGI.eu, "EGI | EGI Foundation," [Online]. Available: https://www.egi.eu/about/egi-foundation/. [Accessed 04 March 2020].
3. EOSC-hub, "About us | EOSC Hub," [Online]. Available: https://www.eosc-hub.eu/about-us. [Accessed 04 March 2020].
4. WLCG, "Welcome to the Worldwide LHC Computing Grid | WLCG," [Online]. Available: https://wlcg.web.cern.ch/. [Accessed 04 March 2020].
5. European Commission, "Enabling Grids for E-sciencE-II | EGEE-II Project | FP6 | CORDIS | European Commission," 31 March 2017. [Online]. Available: https://cordis.europa.eu/project/id/031688. [Accessed 04 March 2020].
6. European Commission, "Enabling Grids for E-sciencE III | EGEE-III Project | FP7 | CORDIS | European Commission," 16 July 2019. [Online]. Available: https://cordis.europa.eu/project/id/222667. [Accessed 04 March 2020].
7. European Commission, "European Grid Initiative: Integrated Sustainable Pan-European Infrastructure for Researchers in Europe | EGI-InSPIRE Project | FP7 | CORDIS | European Commission," 01 August 2019. [Online]. Available: https://cordis.europa.eu/project/id/261323. [Accessed 04 March 2020].

8. European Commission, "ngaging the EGI Community towards an Open Science Commons | EGI-Engage Project | H2020 | CORDIS | European Commission," 04 August 2017. [Online]. Available: https://cordis.europa.eu/project/id/654142. [Accessed 04 March 2020].

9. EGI Foundation, "Glossary V1," [Online]. Available: https://wiki.egi.eu/wiki/Glossary_V1.

10. EGI Foundation, "GOCDB/PI/Technical Documentation," 12 February 2016. [Online]. Available: https://wiki.egi.eu/wiki/GOCDB/PI/Technical_Documentation. [Accessed 27 february 2020].

11. EGI Foundation, "GOCDB/Write API/Technical Documentation," 6 February 2020. [Online]. Available: https://wiki.egi.eu/wiki/GOCDB/Write_API/Technical_Documentation. [Accessed 27 February 2020].

12. European Open Science Cloud, "EOSC," 2018. [Online]. Available: https://www.eosc-portal.eu/about/eosc. [Accessed 27 February 2020].

13. EGI.eu, "APEL - EGIWiki," 27 March 2018. [Online]. Available: https://wiki.egi.eu/wiki/APEL. [Accessed 4 March 2020].

14. ELIXIR, "About us | ELIXIR," [Online]. Available: https://elixir-europe.org/about-us. [Accessed 04 March 2020].

15. GridPP, "RAL Tier1 - GridPP Wiki," 15 December 2019. [Online]. Available: https://www.gridpp.ac.uk/wiki/RAL_Tier1. [Accessed 04 March 2020].

16. Red Hat, "Red Hat Enterprise Linux Life Cycle," [Online]. Available: https://access.redhat.com/support/policy/updates/errata/#Maintenance_Support_2_Phase. [Accessed 12 06 2020].