Grid Information Systems: Past, Present and Future

Laurence Field[1], *Markus* W. Schulz[1], and *Maria* Alandes Pradillo[1]

[1]CERN, Geneva, Switzerland

Abstract. Grid information systems enable the discovery of resources in a grid computing infrastructure and provide further information about their structure and state. The original concepts for a grid information system were defined over 20 years ago and the GLUE 2.0 information model specification was published 10 years ago. This contribution describes the current status and highlights the changes over the years. It provides an overview of today's usage from analysing the system logs and compares this with results from over a decade ago. A critical analysis of the system is provided with lessons learned and some perspectives for the future.

1 Introduction

The CHEP 2019 conference represents the 20 year milestone since the introduction of grid computing to HEP. During the CHEP 2000 conference, discussions took place on the emerging field of grid computing and it was realised that it could potentially provide the technical implementation of the LHC distributed computing environment [1]. Grid computing addresses the challenge of integrating resources across multiple administrative domains through the provision of an authentication and authorisation infrastructure to allow users to seamlessly access resources distributed around the world. Following the CHEP 2000 conference, a series of development projects were launched to prototype grid computing for the LHC and all were based on the Globus toolkit [2] from the Globus project.

In addition to providing the Grid Security Infrastructure (GSI) [3] for authentication and authorisation, the Globus toolkit provided four basic protocols for integrating computing and storage resources. The Grid Resource Access and Management (GRAM) and Grid File Transfer Protocol(GridFTP) protocols were the abstract interfaces for compute and storage respectively. These could be queried using the Grid Resource Information Protocol (GRIP) and discovered as the GRIP endpoints were registered to indexes using the Grid Resource Registration Protocol (GRRP) [4]. Together the GRIP and GRRP protocols provided the building blocks for the Metacomputing Directory Service (MDS) [5], a grid information system implementation.

This paper describes the current status of the grid information system and highlights the changes that have occurred over the years. The current usage is presented and compared to an identical study [6] made 10 years previously. The following section provides a short overview of the grid information systems and its evolution. Section 3 describes the current usage and then compares the results to the previous study. This is discussed in Section 4 along with some perspectives on the future.

2 The Evolution Of The Grid Information System

Grid information systems support coordinated resource-sharing and problem-solving as virtual organisations need to obtain information about the structure and state of grid services which are widely distributed geographically. The information describing a grid service is provided by the service itself hence the grid service is the primary information source. The information provided conforms to an information model to ensure a consistent interpretation of the values across the different implementations and instances of the services. Queries may consider thousands of information sources in order to enable efficient grid functions that may utilise multiple cooperating services. The goal is to efficiently execute many queries, from many clients, across many information sources.

2.1 The Metacomputing Directory Service (MDS)

The MDS from the Globus project provides an implementation of the two information protocols (GRIP and GRRP) and offers a clear separation between inquiry (data retrieval) and discovery (information retrieval). The MDS consists of two basic elements, information providers and information indexing services, which together can be used to build a hierarchy of query processing engines that can forward sub-queries to individual providers and merge the results (query shipping approach). Information providers obtain information about the grid service via local operations and the results are structured to conform to a predefined information model. To control the intrusiveness of queries triggering the execution of information providers and to improve the response time, each information provider's results may be cached for a configurable period of time. Indexing services, collect, manage, index, and respond to information provided by one or more information providers, and are analogous to the catalogue in a typical query processing architecture [7]. The MDS implementation adopted the standard Lightweight Directory Access Protocol (LDAP) [8] for the GRRP transport mechanism, with GRRP messages mapped onto *LDAP add* operations of the standard LDAP protocol, and for the GRIP, where it is used to define the data model, query language, and transport protocol. This choice of LDAP was made for pragmatic reasons in order to simplify development and it was thought at the time that due to the widespread use of the LDAP data formats would be familiar to developers and system administrators [5].

> Not only is the LDAP data representation extensible and flexible, but LDAP is beginning to play a significant role in Web-based systems. Hence, we can expect wide deployment of LDAP information services, familiarity with LDAP data formats and programming, and the existence of LDAP directories with useful information.

Today over 20 years since this prediction was made, one of the criticisms of the grid information system is that the LDAP model, on which it is based, is not familiar and hence is seen as complicated.

The evaluation of MDS [9] during the EDG project [10], which ran from 2001 to 2004, revealed instabilities that contributed to an unacceptably high failure rate for the end-users and frequent interventions were necessary to keep the testbed operational. To work around the issues, a cache based on the standard OpenLDAP server was created by periodically adding the information extracted from the MDS. Queries would subsequently be directed to the OpenLDAP server instead of the MDS. This cache approach continued to be responsive under high query loads and in December 2002 became a standard component of the EDG Middleware. The component was named the Berkeley Database Information Index (BDII)

[11] and was enhanced over the years to completely replace all the MDS components. One of the main differences with this approach was the use of a static list of LDAP and File URLs for information discovery. It also avoided the problems associated with the dynamic nature of the GRRP protocol, in addition it enabled the infrastructure to be centrally managed which was a desirable feature. As the cache was updated asynchronously from the query, it was ensured that the cache was always available and that there was never any significant increase in the query response time. Decoupling the query from the cache update also ensured that the queries would never be propagated to lower levels of the information system hierarchy. These changes resulted in a data shipping approach being adopted whereby the data is obtained from the lower level by querying for all information.

2.2 Information Models

Information models ensure the agreement on the meaning of information. They describe the real entities, the relationships between those entities and their semantics. A data model defines the syntax by which information is exchanged. The original MDS information model described the physical and logical components of a compute resource. The EDG project discovered that the MDS information model did not describe grid entities in sufficient detail for use with higher-level services. The computing resources were managed by batch systems and for higher-level scheduling it was important to know the state of the batch system rather than the state of the individual hosts in the cluster [10]. Similarly for data management, a description of the storage system was required. As such, each area defined its own information model to describe the entities required to enable the higher-level functionalities.

A collaborative effort between EDG and other projects aimed to define a common information model as a step towards transatlantic testbed interoperability. The result was the Grid Laboratory Uniform Environment (GLUE) information model (and LDAP data model) for a uniform representation of Grid resources. A minor revision (v1.2) was undertaken in 2005 to address a number of minor problems with the original version and additional use cases from the Open Science Grid (OSG)[12], a U.S.consortium building and operating a grid infrastructure whose capabilities and schedule were driven by U.S. participants that used the LHC. As such the OSG switched to the GLUE information model and achieved interoperability between the US and LCG infrastructure and this milestone lead to the name Worldwide LHC Computing Grid (WLCG). A year later work began on version 1.3 with the main focus of adding support for the Storage Resource Manager [13] in time for the start-up of the LHC.

By 2007 the majority of grid infrastructures were using the GLUE 1.3 information model [14]. One exception was NorduGrid [15] who were using their own information model, developed around the same time as the EDG model, for the same reasons. The GLUE Working Group was created within the Open Grid Forum to oversee a major revision of the GLUE information model (GLUE v2.0) which would consolidate the NorduGrid and GLUE 1.3 information models into a community agreed standard. As the endeavour built upon existing information models, it benefited from several years of experience in the context of production grid infrastructures. Before in-depth work could be undertaken on information modelling, a number of fundamental grid concepts needed to be agreed. The production of the model took 347 days including 45 phone conferences, 3 days of face-to-face talks and approximately two months of full time man power. Over 40 iterations of the document [16] were made and the result was 46 pages with 12787 words to describe 254 attributes spanning 28 objects. At the core of this model is the concept of a *Service* and as such this demonstrates that a grid has a service-oriented architecture rather than protocols. A *Service* enables a *User* from *User Domain* (VO) to run an *Activity* via an *Endpoint* on a *Share* of a *Resource* in accordance to an *Access Policy*. A Service is therefore a container which is used to describe collection

of sub-components that are required to fulfill a particular function. Services are related to an *Admin Domain* (Computing Site) which provides and manages the Service. For many use cases it is necessary to define a more detailed information model about specific services. Official renderings have been provided for XML, JSON and LDIF to support alternative or future systems.

One interesting to point to note is that 10 years after GLUE 2.0 was defined, GLUE 1.3 information is still being published and consumed. This suggests that the GLUE 1.3 information model was sufficient and there were no critical use cases that could only be met with GLUE 2.0, hence little pressure to transition. As the main motivation for GLUE 2.0 was to create a community agreed standard and to emphimprove the model from a conceptual point of view, this was more relevant for those who were not already using GLUE 1.3. Interoperation between different grid infrastructures was therefore the main driver for its adoption and use.

2.3 Information Validation

An assumption with the information system is that the information source is up-to-date, that is the values represent the real state of the grid service. While information models go a long way to ensure information validity, however conformance checks have their limitations. For example:

- Is information missing or non-existent?

- Does the state published actually reflect the state of the system?

- Are values published for the correct units e.g. bytes vs gigabytes

These examples are not exhaustive but serve to highlight some of the issues of information systems. It is important to note that the information sources are distributed and information providers are critical components that interact with the system to extract and form the information in accordance with the model. Therefore as many checks as possible should be done at the information provider level before the information is published.

3 Evolution Of The Information System Content and Usage

In the paper titled *Scalability and performance analysis of the EGEE information system* [6], the content and usage of the information was described. At the time (CHEP 2007), the grid was comprised of 251 sites which provided 1428 Services and the information system (lcg-bdii.cern.ch) handled 2 million connections per day. Today it handles 1 million queries per day and there are 209 sites providing 883 Services (GLUE 2.0) or 200 sites providing 909 Services (GLUE 1.3). These values alone do not provide the full picture for the period, however since March 2010 daily snapshots of the information system have been archived and can be used for historical studies. Figure 1 shows the number of sites and services seen in the grid information system since 2010. The peak number of sites was 389 in 2012 and the number of services has been steadily decreasing during that period. The acceleration of services decreasing during 2014 corresponds to the decision by OSG to stop publishing information with the last OSG site leaving in 2015. Although the number of sites and services have been decreasing, the number of cores and the storage available continues to increase.

The paper [6] also documented the top ten queries found from analysing the usage logs and these can be seen in Table 1. This exercise has been repeated and the result can be seen in Table 2. One major difference is that in 2007 six of the queries were related to storage whereas in 2019 only one was storage specific. One reason for this change is that
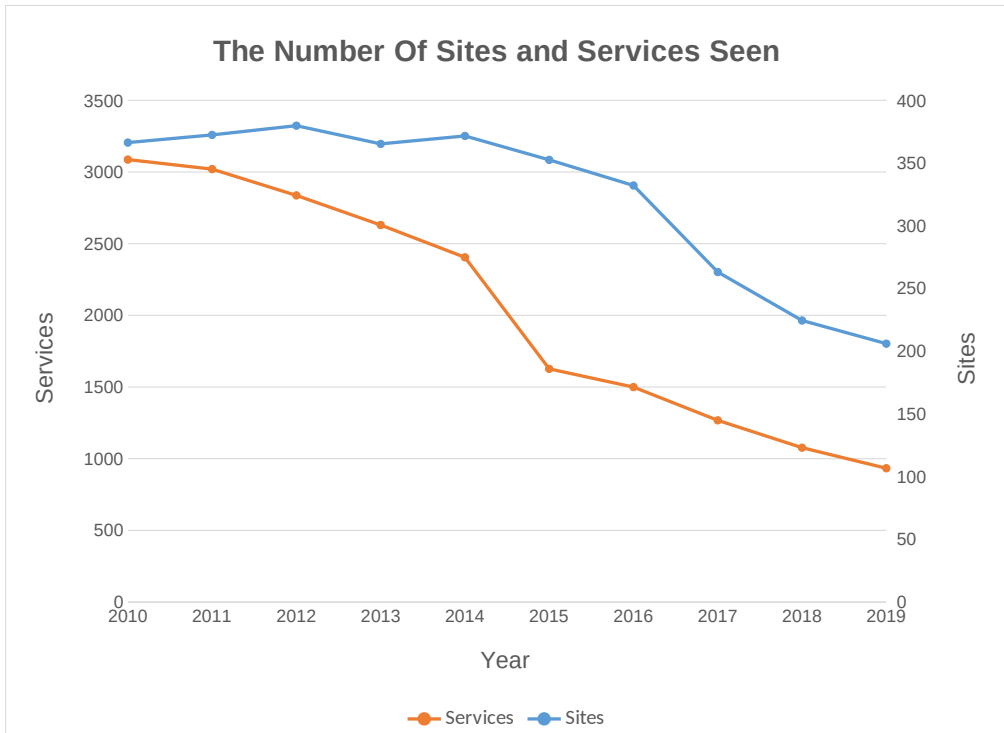
Figure 1. The number of sites and services seen in the grid information system (GLUE 1.3).

Q/h	Query
6075	Close CE to an SE
5475	VO's SA for an SE
5043	All SRMs
4791	An SE
2432	Close SE to a CE
2117	All Services for a VO
664	All CEs for a VO
638	All SAs for a VO
479	All SubClusters
448	GlueVOViewfor a CE

Table 1. Top ten queries from 2007

Q/h	Query
5960	A specific Cluster
5923	All entries linked to a Cluster
5377	*EEs of a Cluster*
4898	*GLUE2Shares for a VO*
2928	A specific Site
909	SRM endpoint of a SE
305	Find all CEs for a VO
217	Find a specific CE for a VO
193	*A specific GLUE2share*
134	Cream CEs for a VO

Table 2. Top ten queries from 2019. GLUE 2.0 queries in italics.

the workload management systems stopped attempting to do global resource brokering and moved towards the pilot job paradigm [17]. It can also be seen that although GLUE 2.0 was defined 10 years ago, many queries are still searching for GLUE 1.3 information.

4 Current Status and Future Plans

From Figure 1 it appears that the grid infrastructure is shrinking in terms of participation, however Tables 1 and 2 show that although the load on the information system is similar, the nature of the queries is changing. One question that needs answering is how relevant is the information system today? An answer can be found in the commissioning of a new grid service implementation, the HTCondor CE [18]. The HTCondor CE is a new grid interface to the HTCondor [19] batch system, similar in scope to GRAM. It has been adopted as the grid interface for the CERN batch system to replace the functionality of the CREAM CEs [20]. As part of the deployment a new provider was required to be produced for HTCondor. Since this was new, a decision was made to only publish the information using the GLUE 2.0 information model. It was developed by initially only providing the minimal attributes that were required and extended based on requests for additional information. From this we can conclude that the information is required as there were specific requests for it. It is interesting to note that there were no request for GLUE 1.3 information. The information provider was included in the upstream release of the HTCondor CE and it has since been adopted by other sites.

As for the future, the options are the same as presented in 2011 [21]; Lazy, Radical, Slow and Steady or Rocky. The Lazy option is to do nothing and leave the system running in the current state. This is the default option and has been the situation for the past few years. The Radical option is to decommission the system. If this option is chosen, the process should be actively managed. Usage logs are available which show the queries from the different sources. These can be used to understand the use case for the query, if it is still valid and whether there is an alternative to provide the information required. The Slow and Steady option is to make small changes over time to improve and simplify the system. Examples would be to drop GLUE 1.3 information and to streamline GLUE 2.0 usage along with redundant aspects of the information system. The Rocky option has two separate the use cases. Provide a centralised service discovery system along with a single system for experiment annotation and configuration. This is currently a direction that is already underway with the development of the Computing Resource Information Catalog (CRIC) service [22].

5 Conclusion

The CHEP 2019 conference represents the 20 year milestone since the beginning of the adoption of grid computing for the LHC computing environment. During that period the grid information system has transitioned from an initial prototype to a production infrastructure that is in continuous operation for WLCG. 2019 was also the 10 year anniversary of the publication of the GLUE 2.0 information model. After an initial period of growth, the past five years has seen a decline in the number of sites and services published since a peak of 389 sites in 2012. Today the experience from deploying the HTCondor CE at CERN suggests that the information system is still relevant and necessary. However, it also suggests that the GLUE 1.3 information model can be retired.

Whatever path is taken in the future, the challenges remain the same. In a distributed computing environment such as WLCG, there is the need to discover what services exist and further information about their structure and state. To ensure the agreement on the meaning of the information shared, an information model is required to describe the real entities, the relationships between those entities and their semantics. Information providers are required to interact with the local system to extract the required information and format it in accordance with the information model. The importance of information providers should not be underestimated as their correctness will influence the overall quality of the system.

The current system is based on an short-term workaround that became an official component in 2002. It will be interesting to see how much longer it will last.

References

[1] I. Bird, Annual Review of Nuclear and Particle Science **61**, 99 (2011)

[2] I. Foster, C. Kesselman, International Journal of High Performance Computing Applications **11**, 115 (1997)

[3] I. Foster, C. Kesselman, G. Tsudik, S. Tuecke, *A security architecture for computational grids*, in *Proceedings of the 5th ACM conference on Computer and communications security* (ACM Press, San Francisco, CA, USA, 1998), pp. 83–92, ISBN 1581130074, `http://portal.acm.org/citation.cfm?doid=288090.288111`

[4] K. Czajkowski, S. Fitzgerald, I. Foster, C. Kesselman, *Grid information services for distributed resource sharing*, in *Proceedings 10th IEEE International Symposium on High Performance Distributed Computing* (San Francisco, CA, USA, 2001), pp. 181–194, `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=945188`

[5] S. Fitzgerald, I. Foster, C. Kesselman, G. von Laszewski, W. Smith, S. Tuecke, *A directory service for configuring high-performance distributed computations*, in *Proceedings of the Sixth IEEE International Symposium on High Performance Distributed Computing* (Portland, OR, USA, 1997), pp. 365–375, `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=626445`

[6] F. Ehm, L. Field, M.W. Schulz, Journal of Physics: Conference Series **119** (2008)

[7] R. Baeza-Yates, B. Ribeiro-Neto, *Modern information retrieval* (Addison Wesley, Harlow, England, 1999), ISBN 9780201398298

[8] J. Sermersheim, *Lightweight Directory Access Protocol (LDAP): The Protocol* (2006), `http://tools.ietf.org/html/rfc4511`

[9] C. Loomis, EU Deliverable DataGrid-06-D6.8-414712-3-0 (2003), `https://edms.cern.ch/file/414712/3.0/D6.8-3.0.pdf`

[10] B. Segal, L. Robertson, F. Gagliardi, F. Carminati, *Grid computing: the European DataGrid project*, in *IEEE Nuclear Science Symposium Conference Record* (Lyon, France, 2000), Vol. 1, `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=948988`

[11] Laurence Field, Markus Schulz, *Grid deployment experiences: evolution of the information and monitoring system*, in *Proceedings of the Conference for Computing in High-Energy and Nuclear Physics* (Mumbai, India, 2006)

[12] R. Pordes, D. Petravick, B. Kramer, D. Olson, M. Livny, A. Roy, P. Avery, K. Blackburn, T. Wenaus, F. Würthwein et al., Journal of Physics: Conference Series **78** (2007)

[13] F. Donno, L. Abadie, P. Badino, J.P. Baud, E. Corso, S.D. Witt, P. Fuhrmann, J. Gu, B. Koblitz, S. Lemaitre et al., Journal of Physics: Conference Series **119** (2008)

[14] Morris Riedel, Concurrency and Computation: Practice and Experience **21**, 961 (2009)

[15] M. Ellert, A. Konstantinov, B. Kónya, O. Smirnova, A. Wäänänen, *The NorduGrid project: using Globus toolkit for building grid infrastructure*, in *Proceedings of the VIII International Workshop on Advanced Computing and Analysis Techniques in Physics Research* (Moscow, 2002), Vol. 502, pp. 407–410, `http://linkinghub.elsevier.com/retrieve/pii/S0168900203004534`

[16] L. Field, B. Konya, S. Andreozzi, Tech. Rep. GFD.147, Open Grid Forum (2009), `http://www.ogf.org/documents/GFD.147.pdf`

[17] S. Paterson, R. Graciani, A. Tsaregorodtsev, A. Casajus, Journal of Physics: Conference Series **219**, 062049 (2010)

[18] B. Bockelman, T. Cartwright, J. Frey, E.M. Fajardo, B. Lin, M. Selmeci, T. Tannenbaum, M. Zvada, Journal of Physics: Conference Series **664**, 062003 (2015)

[19] M. Litzkow, M. Livny, M. Mutka, *Condor-a hunter of idle workstations*, in *Proceedings. The 8th International Conference on Distributed* (San Jose, CA, USA, 1988), pp. 104–111, `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=12507`

[20] P. Andreetto, S. A. Borgia, A. Dorigo, A. Gianelle, M. Marzolla, M. Mordacchini, M. Sgaravatto, S. Andreozzi, M. Cecchi, V. Ciaschini et al., *Cream: A simple, grid-accessible, job management system for local computational resources*, in *In Proceedings of CHEP'06, Mumbai* (Mumbai, 2006)

[21] Lorenzo Dini, *Information System Use Cases and Future Steps*, `https://indico.cern.ch/event/106645/contributions/1306710/attachments/20850/30147/Information_System_Use_cases_2011_06_08_v2.pdf`

[22] A. Anisenkov, J. Andreeva, A. Di Girolamo, P. Paparrigopoulos, A. Vedaee, EPJ Web of Conferences **214**, 03003 (2019)