

EGI Dataset Accounting and the WLCG

Tuesday 10 July 2018 11:00 (15 minutes)

While the WLCG and EGI have both made significant progress towards solutions for storage space accounting, one area that is still quite exploratory is that of dataset accounting. This type of accounting would enable resource centre and research community administrators to report on dataset usage to the data owners, data providers, and funding agencies. Eventually decisions could be made about the location and storage of data sets to make more efficient use of the infrastructure. By giving insight to data usage, dataset accounting also assists scientists in assessing the impact of their work.

This paper reviews the status of the prototype dataset accounting developed during EGI-Engage and how it could be used to complement the view that the WLCG has of its datasets. This is a new feature of the EGI resource accounting system that will enable storing information on dataset usage such as who has accessed a dataset and how often, the transfer volumes and end points etc. The design of this new feature has been led by the users' requirements collected in the first part of the project, from which a set of dataset accounting metrics were derived. In these trials, the EGI Accounting Repository was integrated with the data provider Onedata (the underlying technology powering the EGI Open Data Platform and EGI DataHub) as an example of a generic data provider.

Primary authors: Mr COVENEY, Adrian (STFC); Mr CORBETT, Greg (STFC); COLLIER, Ian (STFC-Rutherford Appleton Laboratory (GB))

Presenter: Mr COVENEY, Adrian (STFC)

Session Classification: T3 - Distributed computing

Track Classification: Track 3 –Distributed computing