

EGU2020-18310

<https://doi.org/10.5194/egusphere-egu2020-18310>

EGU General Assembly 2020

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Evolution and Future Architecture for the Earth System Grid Federation

Philip Kershaw¹, Ghaleb Abdulla², Sasha Ames², Ben Evans³, Tom Landry⁴, Michael Lautenschlager⁵, Venkatramani Balaji^{6,7}, and Guillaume Levvasseur⁷

¹STFC Rutherford Appleton Laboratory, RAL Space, Didcot, United Kingdom of Great Britain and Northern Ireland (philip.kershaw@stfc.ac.uk)

²LLNL, Livermore, USA

³NCI, Australian National University, Acton, Australia

⁴CRIM, Montréal, Canada

⁵DKRZ, Hamburg, Germany

⁶Princeton University, Princeton, USA

⁷IPSL, Paris, France

The Earth System Grid Federation (ESGF) is a globally distributed e-infrastructure for the hosting and dissemination of climate-related data. ESGF was originally developed to support the community in the analysis of CMIP5 (5th Coupled Model Intercomparison Project) data in support of the 5th Assessment report made by the IPCC (Intergovernmental Panel on Climate Change). Recognising the challenge of the large volumes of data concerned and the international nature of the work, a federated system was developed linking together a network of collaborating data providers around the world. This enables users to discover, download and access data through a single unified system such that they can seamlessly pull data from these multiple hosting centres via a common set of APIs. ESGF has grown to support over 16000 registered users and besides the CMIPs, supports a range of other projects such as the Energy Exascale Earth System Model, Obs4MIPS, CORDEX and the European Space Agency's Climate Change Initiative Open Data Portal.

Over the course of its evolution, ESGF has pioneered technologies and operational practice for distributed systems including solutions for federated search, metadata modelling and capture, identity management and large scale replication of data. Now in its tenth year of operation, a major review of the system architecture is underway. For this next generation system, we will be drawing from our experience and lessons learnt running an operational e-infrastructure but also considering other similar systems and initiatives. These include for example, ESA's Earth Observation Exploitation Platform Common Architecture, outputs from recent OGC Testbeds and Pangeo (<https://pangeo.io/>), a community and software stack for the geosciences. Drawing from our own recent pilot work, we look at the role of cloud computing with its impact on deployment practice and hosting architecture but also new paradigms for massively parallel data storage and access, such as object store. The cloud also offers a potential point of entry for scientists without access to large-scale computing, analysis, and network resources. As trusted international

repositories, the major national computing centres that host and replicate large corpuses of ESGF have increasingly been supporting a broader range of domains and communities in the Earth sciences. We explore the critical role of standards for connecting data and the application of FAIR data principles to ensure free and open access and interoperability with other similar systems in the Earth Sciences.