EGU22-12105
https://doi.org/10.5194/egusphere-egu22-12105
EGU General Assembly 2022

User Identification and Authentication for Geophysical Data Centers: Exploring a Difficult Transition

Florian Haslinger[1], Jerry Carter[2], Helle Pedersen[3], Jonathan Schaeffer[4], Robert Casey[2], Javier Quinteros[5], and Angelo Strollo[5]

[1]ETH Zürich, Schweizerischer Erdbebendienst, Zürich, Switzerland (haslinger@sed.ethz.ch)
[2]IRIS, USA
[3]University of Grenoble Alpes, University of Savoie Mont Blanc, CNRS, IRD, University of Gustave Eiffel, ISTerre, Grenoble, France
[4]University of Grenoble Alpes, Irstea, CNRS, IR, OSUG, Grenoble, France
[5]Deutsches GeoForschungsZentrum (GFZ), Potsdam, Germany

Many geophysical data centers are being asked by their sponsors and funding agencies to provide information on what data and services are used by whom and for what purpose in greater detail than customary in the past, when bulk information about the number of users/accesses and volumes of download were deemed sufficient in most cases. Up to now, data centers generally offer anonymous access to large parts of their holdings, with different approaches to basic monitoring and access logging, e.g. by IP address, as a rough proxy, that allows one to infer geographical user distribution to some detail.

Already today, access to embargoed or otherwise restricted data, or to advanced functions like personal work spaces and computational resources, is usually protected by user authentication and authorisation. Standardization of the identity management protocols is a requirement for further supporting the federation of data centers and their services, also in light of future integration with cloud services or other integrated services. For example in seismology, federated data retrieval systems follow a specific credential process based on standards for data exchange and web services established and maintained by the International Federation of Digital Seismograph Networks (FDSN).

These new information requirements from funding agencies would, however, require implementing identity management systems and some sort of user identification / authentication to many or all data center services and resources. This is raising concerns within the data centers on a number of aspects: Evidence from other domains demonstrates that requiring authentication reduces the use of data center services; enforcing authentication is often perceived as being not in line with best practices for open science; implementing identity management for usage profiling may lead to significantly increased effort at the data centers, especially with regard to compliance with data protection legislation like GDPR, and it may significantly impede automated (scripted) machine-to-machine access; the level of detail that should be reported back to funding agencies is unclear and there are doubts whether detailed user profiling is a reasonable ‘performance

indicator'. Indeed, such knowledge gathering on users needs to be obtained through technical implementations that take into account the impact on user experience, the impact on decades of research tool development, and the resources necessary to implement and operate such systems, whether embedded into the operational services or taking other forms such as surveys and outreach to user groups.

Relevant discussions have now started among representatives of major geophysical data centers so that interim plans can be shared, ideas and experiences exchanged, and standard approaches can be developed and recommended for consideration by the community. In these discussions we consider both scenarios where identity management is useful and relevant or where we may consolidate our views and arguments with respect to the general user data reporting requests.