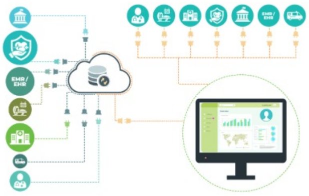
HealthManagement
Promoting Management and Leadership

Interoperability of Cloud Platforms: FAIR Data in SAFE Environments



As the number of cloud platforms supporting scientific research grows, there is an increasing need to support cross-platform interoperability. With interoperability between cloud platforms, data does not have to be replicated on multiple cloud platforms but can be managed by one cloud platform and analysed by researchers on another cloud platform. Interoperability also enables cross-platform functionality, allowing researchers analysing data on one cloud platform to obtain the necessary amount of data required to power a statistical analysis on another cloud platform. In a recent study published in *Scientific Data*, we are especially concerned with frameworks that are designed to support the interoperability of sensitive or controlled access data, such as biomedical data or qualitative research data.

Several frameworks for interoperability

Several frameworks have been developed to enable interoperability between cloud platforms for biomedical data, such as those by the GA4GH organisation and the European Open Science Cloud (EOSC) Interoperability Task Force. These frameworks aim to make data findable, accessible, interoperable, and reusable (FAIR). However, challenges arise due to governance structures that often require sensitive data to remain within a single cloud platform, hindering interoperability. Reasons for this include policies against data removal, security and compliance management, and perceived risks to sponsors. To address these issues, a new concept called Secure and Authorized FAIR Environment (SAFE) is introduced. SAFE aims to facilitate interoperability between cloud platforms for restricted-access research data, such as biomedical and social science data. SAFE platforms support governance decisions regarding data linkage and transfer between platforms, without compromising security or compliance. The SAFE framework allows sponsors to extend their boundaries to selected third-party platforms for data analysis by authorised users, thus enabling researchers to use familiar tools and platforms.

Interoperability needs processes for authorising environments

The suggested processes for authorising environments involve reviewing their security and compliance by appropriate officials or committees determined by platform governance. These SAFE environments should have APIs to enable findability, accessibility, and interoperability with other cloud platforms. A SAFE environment is a cloud platform approved through platform governance and exposes an API for interaction with other platforms. Examples of functionality to be exposed by the API and proposed identifiers are discussed, focusing on attestation and approvals to support interoperability. The framework applies to all controlled-access data types, with decisions on authorised users and platforms depending on data sensitivity. SAFE utilises identifiers such as Authorized Platform Identifier (APID), Authorized Platform Network Identifier (APNI), and Authorized Region ID (ARID) to enable interoperability. Cloud platforms should support APIs exposing metadata, including APID, a list of APNIs, and dataset metadata specifying data distribution rights, approved platform networks, and regional restrictions.

SAFE environments: a complement to FAIR data

Study authors present SAFE environments as a complementary concept to FAIR data, aiming to establish a seamless trust relationship between cloud platforms hosting FAIR data and those designated as SAFE environments for data analysis. It advocates for APIs and identifiers within the SAFE framework to facilitate interoperability, focusing on the need for clear attestation and approval processes to support this interoperability. Focus is on the broad applicability of the SAFE framework to various types of controlled-access data, including clinical, genomic, imaging, and environmental data. Decisions regarding authorised users and platforms are noted to depend on the sensitivity of the data, with more stringent conditions applied to highly sensitive data. Specific identifiers are used within the SAFE framework, such as the Authorized Platform Identifier (APID), Authorized Platform Network Identifier (APNI), and Authorized Region ID (ARID), which are crucial for enabling interoperability between cloud platforms. Some examples of platform governance frameworks exist, including processes followed by organisations like the NIH Data Access Committees and the NIST 800-53 framework, to illustrate how cloud platforms can be approved as authorised environments for hosting controlled-access data. The concept of the "right to distribute" data highlights the need for stringent data governance processes to ensure data is distributed securely and only to authorised users and platforms. In terms of interoperability, the concepts of authorised users, authorised environments, and the right to distribute data are fundamental to achieving interoperability between cloud platforms. Transparent requirements definition by project sponsors is crucial to facilitate interoperability and accelerate research outcomes.

By standardizing the properties to be a SAFE environment and agreeing to the principle that the data governance process for a dataset should authorize users and the platform governance process should authorize cloud platform environments, then all that is required for two or more cloud platforms to interoperate is for the cloud platforms to trust these authorizations. We can shorten this principle to: “authorize the users, authorize the cloud platforms, and trust the authorizations.” This is the core basis for interoperability in the SAFE framework.

Source: Nature

Image Credit: iStock

Published on : Wed, 6 Mar 2024