RNP provides Internet connectivity services throughout the Brazilian territory, serving universities, hospitals, and research institutes. Due to the large size of the network, an enormous volume of monitoring and performance data is generated and stored. To facilitate and even promote the use of this data in scientific research and bring transparency to the services provided, RNP initiated the Network Data Catalog/DataX System. This system aims to facilitate the identification, location, and usage of network data collected from the Brazilian backbone infrastructure. To achieve this objective, this project presents the knowledge of the scientific community and, eventually, meets, in an agile and efficient way, the demands of researchers related to access to network data collected and processed by the institution.

The Network Data Catalog/DataX is a system that integrates two fronts: (1) the creation of an access portal where researchers and members of the scientific community can find this data described at a higher semantic level, facilitating the search for the correct type of data; and (2) integration with the DataX engine, which is a tool that allows RNP to apply predefined access policies, transformations in the raw data and, finally transfer the selected dataset to the researcher.

In our demonstration, we will present how data sharing can be accomplished from two perspectives: the requesting researcher and the data operator. Researchers can access the Network Data Catalog when they need specific datasets for new experiments. The Catalog allows researchers to search for the desired data directly, conducting a broader exploration of the data available in the dataset, possibly abstracting the specific desired data. The Catalog provides mechanisms for researchers to navigate through terms representing more generic and abstract data concepts until reaching terms that describe details of RNP's dataset sources where concrete data can be obtained. Once the dataset is selected, the researcher must request the dataset from RNP, which may use the resources of DataX to extract them from the database.

The DataX system acts as a tool for exporting datasets from the RNP database to the repository of the requesting researcher. The system was implemented using the open-source Spring Cloud Data Flow framework, which implements ways to operate and program data flows and pipelines by integrating various network data sources and applying policies and process representation. First, the system is informed of which subset of data should be exported and the repository to which the data should be sent. Then, the datasets are collected from the RNP dataset and passed through a broker responsible for routing and implementing a sequence of treatments based on the user's access level and sensitivity of each piece of information. These treatments may include data sanitization, removal of sensitive fields, structural transformation, and/or any RNP corporate policy necessary for sharing. Finally, the data is delivered and stored in the repository specified by the requesting researcher.

The objective of this demonstration will be to present to TNC 2024 participants the solution developed in Brazil to offer network data to the scientific community. We believe that this demonstration may in a certain way: (1) inspire the development of similar solutions around the world; and, (ii) eventually, identify eventual collaborations with RNP regarding the sharing of ideas and data to support the international scientific community.

This demo aims to present the Network Data Catalog/DataX, a system that: (1) offers to researchers in Brazil, RNP's networking data description at a higher semantic level, facilitating the search for the correct type of data; and (2) integration with the DataX engine, which is a tool that allows RNP to apply access policies, data transformations and dataset transfer.

Autores:

Vitor Fontana Zanotelli,

Nilson Luís Damasceno,

Arthur Almeida Vianna,

Gustavo Araujo,

Michael Prieto Hernandez,

Giovanni Comarela,

Magnos Martinello,

Antonio A. de A. Rocha

Daniel da Silva Neto

Allex Magno

Marcos Schwarz