RNP new initiative about data sharing

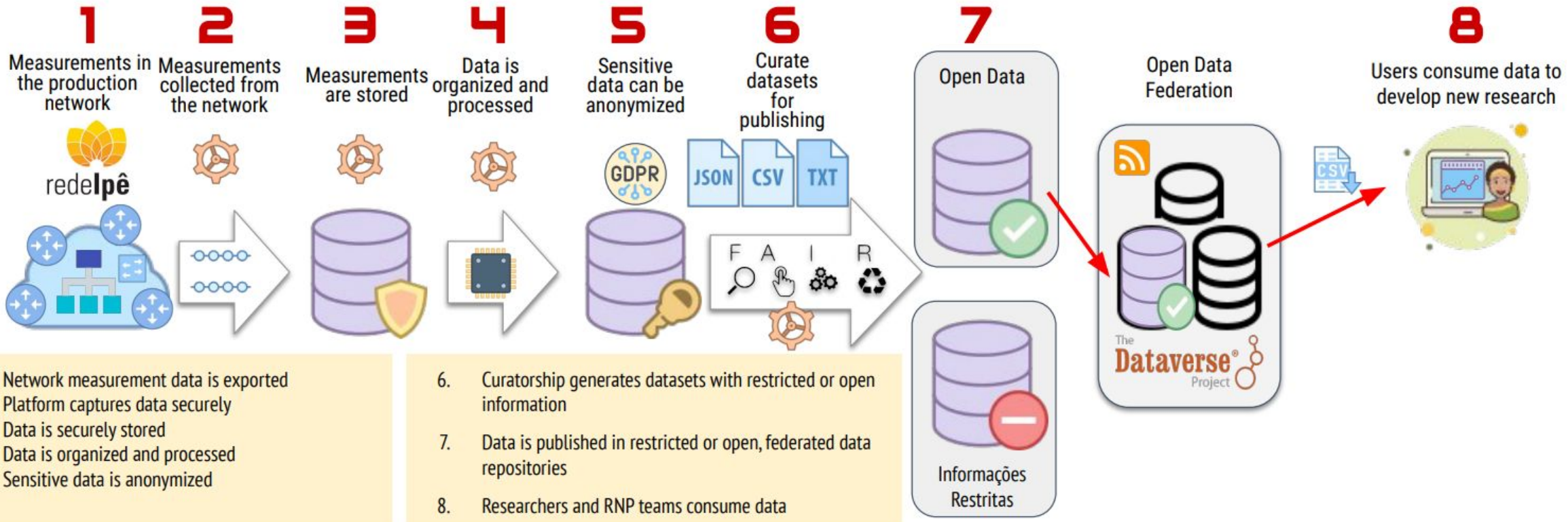
- Marcos Schwarz - R&D Manager
- Michael Prieto Hernandez - R&D Coordinator

2nd Telemetry and Data Workshop

MicroMon Project Overview

Project Objectives: Automated Network Data for Research

- Develop and deploy a new internal process and solution capable of collect, store, organize, anonymize and share - in an automated fashion, OSS data with both the research community and RNP's teams



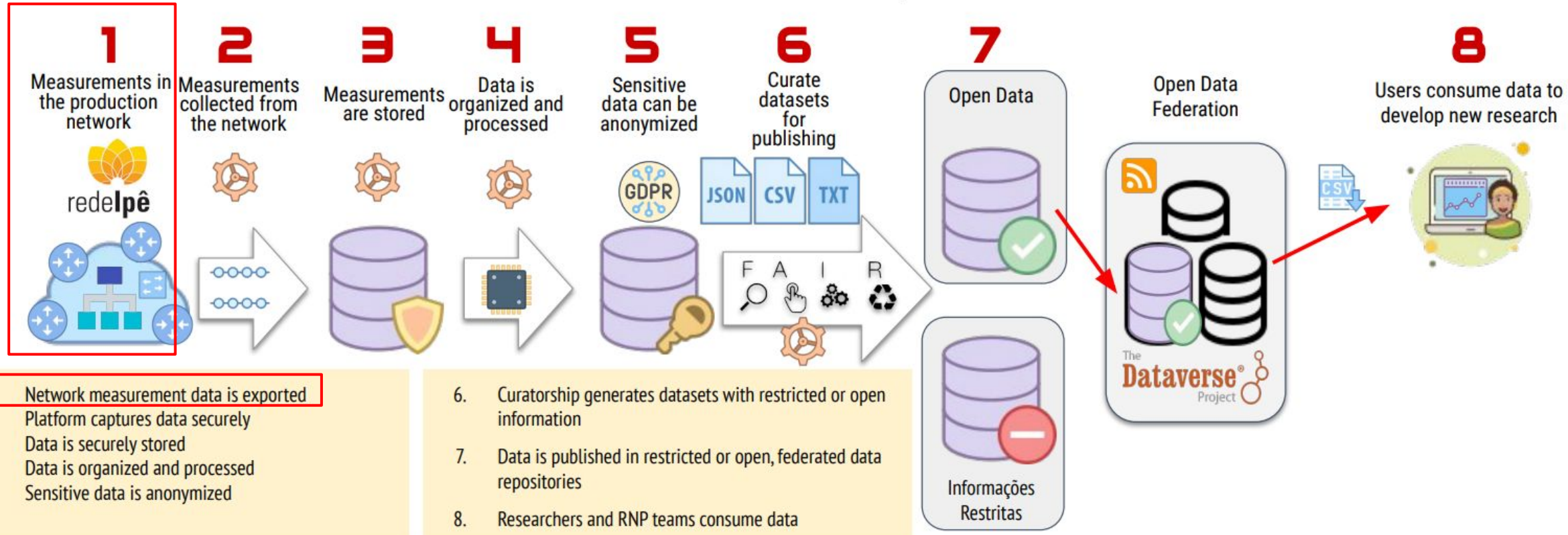
1. Network measurement data is exported
2. Platform captures data securely
3. Data is securely stored
4. Data is organized and processed
5. Sensitive data is anonymized

6. Curatorship generates datasets with restricted or open information
7. Data is published in restricted or open, federated data repositories
8. Researchers and RNP teams consume data

MicroMon Project Overview

Project Objectives: Automated Network Data for Research

- Develop and deploy a new internal process and solution capable of collect, store, organize, anonymize and share - in an automated fashion, OSS data with both the research community and RNP's teams



1. Network measurement data is exported
2. Platform captures data securely
3. Data is securely stored
4. Data is organized and processed
5. Sensitive data is anonymized

6. Curatorship generates datasets with restricted or open information
7. Data is published in restricted or open, federated data repositories
8. Researchers and RNP teams consume data

Challenges and lessons learned

MicroMon was composed of a very capable research team from the academia, and created a great overall solution, but was heavily imparied by the lack of access to data sources

RNP still lacks internal processes, resources and integrations to collect, organize and share data sources of interest for researchers

- Data sharing begins with the Networking Engineering and Operations teams and can’t be done reactively

After 1 year, at August/2021, MicroMon project was stopped with the conclusion that were internal processes missing at RNP to support this initiative

- MicroMon is expected to be resumed in the future

2022 take on data sharing

Internal effort in collaboration between R&D, Network Engineering and Operations

- Focus on sharing existing data used for operations
- Extending existing tools and processes, improving the maintainability
- Generating internal knowledge and pre-incubated, without the need for technology transfers

Identify internal needs for data sharing

- Between networking services/applications, between administrative domains
- Focus on making the required data available for internal use
- But also making it available as a data sample to researchers

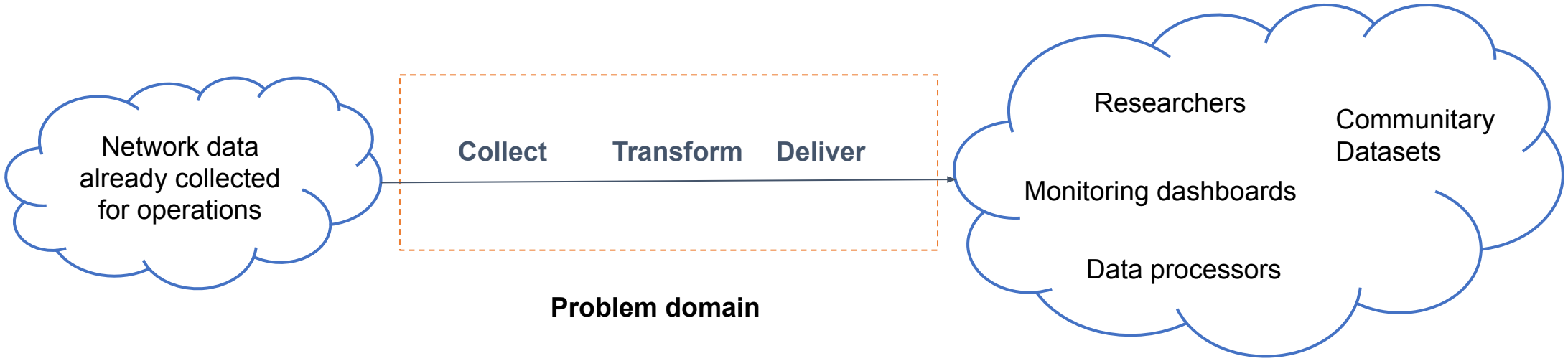
Create a framework to allow the separation of policy enforcement from the data owners

- Provide temporary permits to access sub-sets of a dataset to federated users
- Provide pre-validated anonymization workflows for sensitive data

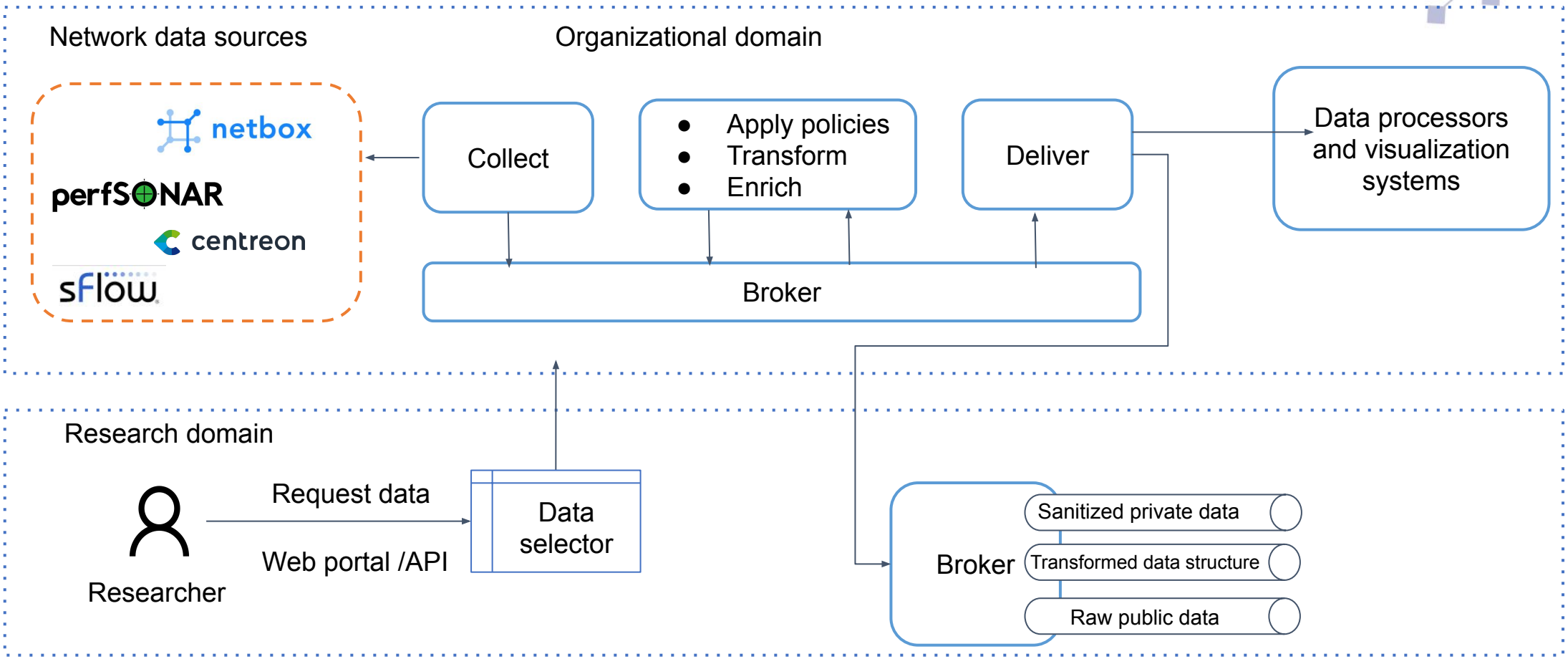
Create an internal process with the Information Management Department to define and implement policies for sharing sensitive date

What are we building:

- A system for sharing network data between different teams inside our organization and with the community
- A system based on a well known simple architecture, but powerful enough to grow over time with different data sources, policies and use cases.



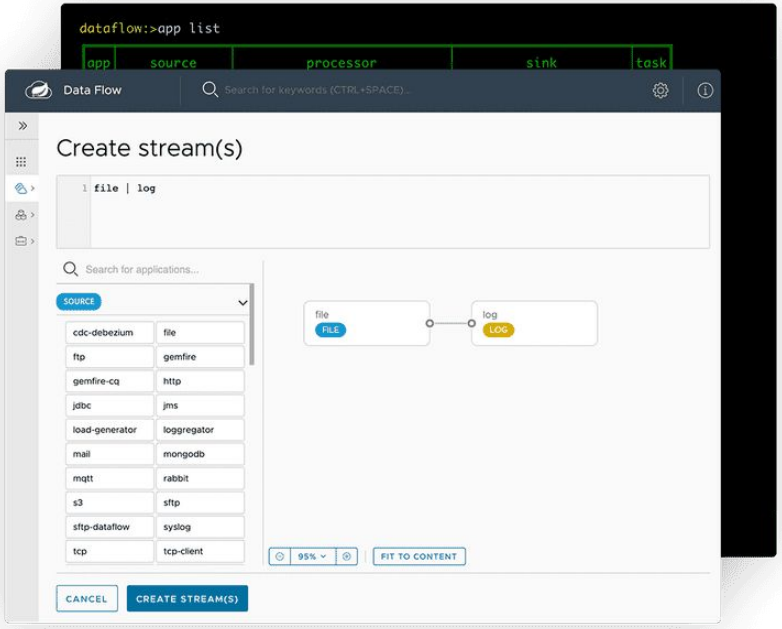
Architecture



Tooling

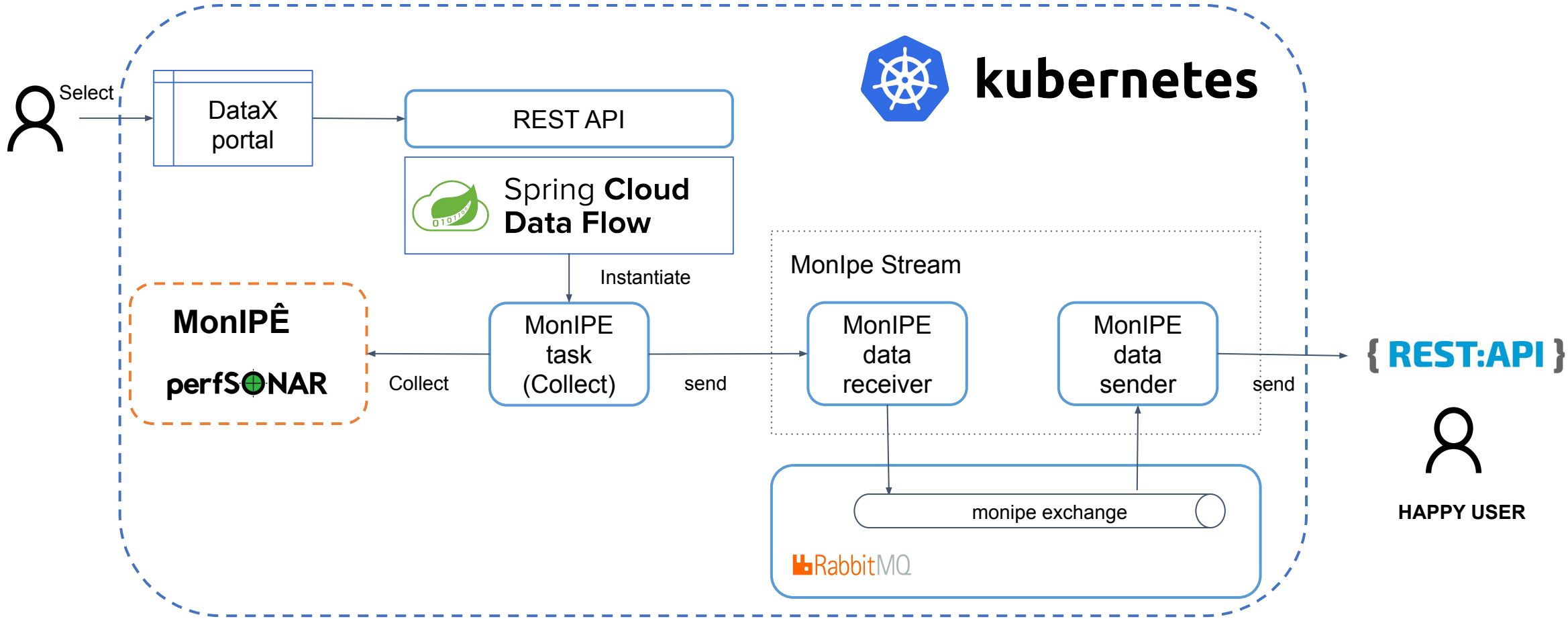
Spring Cloud Data Flow

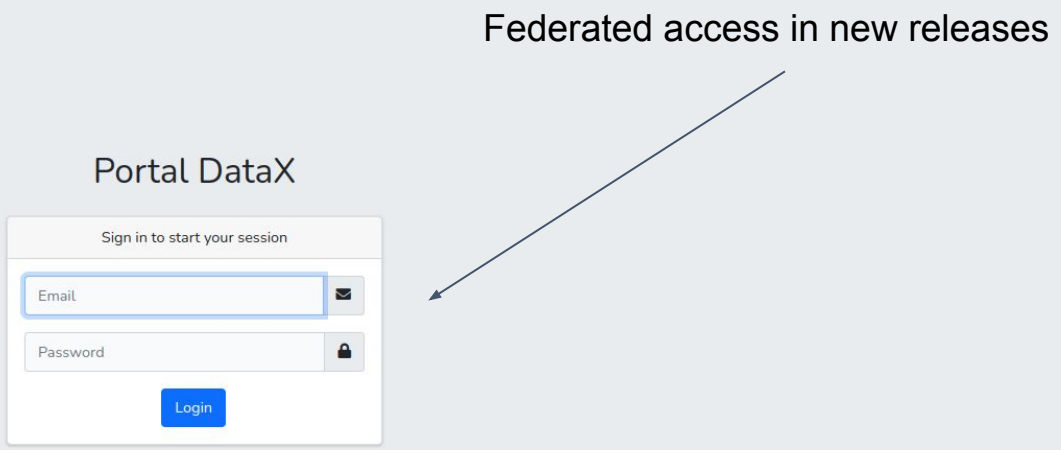
- Provides tools to create complex topologies for streaming and batch data pipelines.
- The data pipelines consist of Spring Boot apps, built using the Spring Cloud Stream or Spring Cloud Task microservice frameworks
- Simplifies the development and deployment of applications that are focused on data-processing use cases.



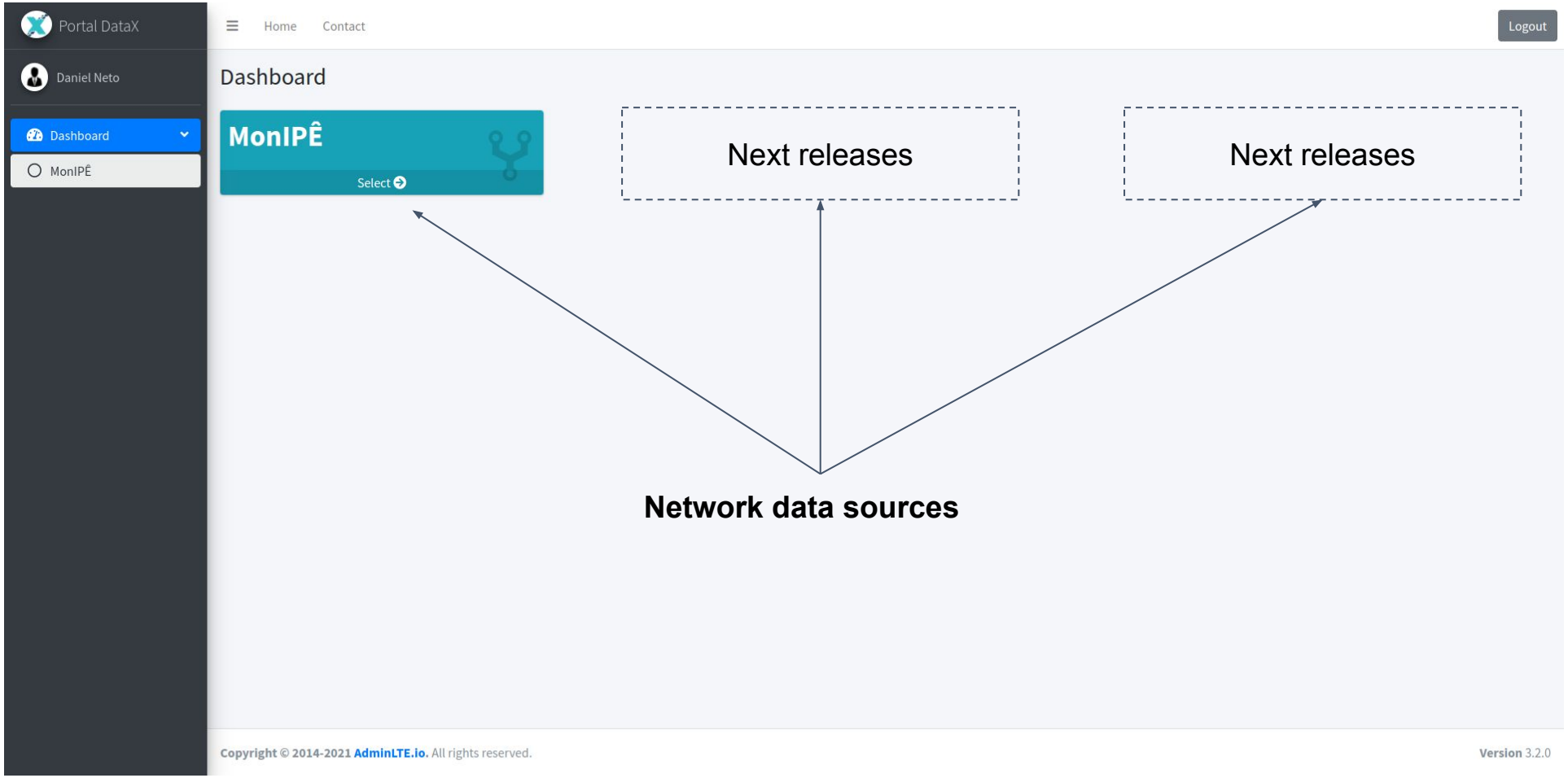
First sprint (Current status): MonIPÊ use case

MonIPÊ: RNP's perfSONAR service

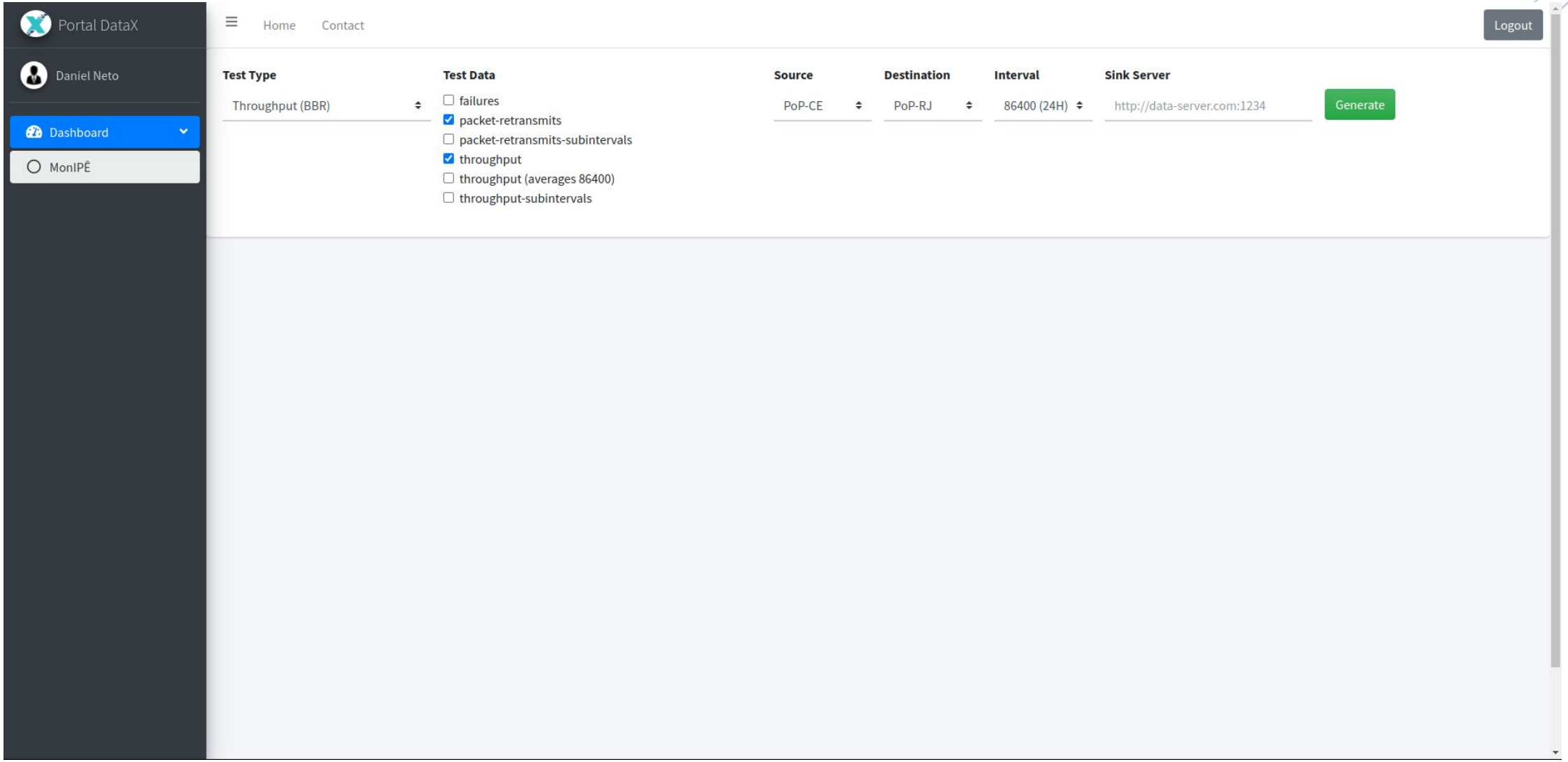



Federated access in new releases
Portal DataX
Sign in to start your session
Email
Password
Login

2nd Telemetry and Data Workshop



2nd Telemetry and Data Workshop



Spring Cloud Data Flow monitoring system

Next steps (Six month)

Use cases:

- Network data from RNP Backbone (exported routers flows with sensitive information)
- Access Circuits performance metrics

Portal

- Federated access
- Scheduled data requests

Thank you