TU Wien: Boris Sedlak, Victor Casamayor Pujol , Praveen Kumar Donta, Schahram Dustdar
{b.sedlak, v.casamayor, pdonta, dustdar}@dsg.tuwien.ac.at
TU Berlin: Sebastian Werner, Karl Wolf, Frank Pallas, Stefan Tai
{sw,kw,fp,st}@ise.tu-berlin.de
POLIMI: Matteo Falconi, Pierluigi Plebani
{matteo.falconi,pierluigi.plebani}@polimi.it

Towards Serverless Data Exchange Within Federations

Problem Definition

- Business data contains confidential or personal information
- Transformations (e.g., privacy enforcement) introduce **data friction**
- Consumers require tailored data sets (e.g., custom data format)
- No platform for **discovering** data sets and **negotiating access**
- Sharing agreements manually negotiated (e.g., healthcare)

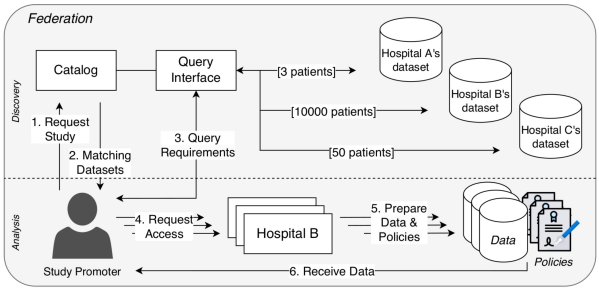

Fig. 1: Ideal data exchange workflow for running joint medical studies

Envisioned Solution

- Data exchanged as **federated data products**
- Customizable data sharing according to policies
- Automatic provisioning of storage / computing resources (e.g., ad hoc or premises of the federation members)
- Automatable negotiation of agreements to protect interests
- **For data providers, alleviate the burden of data sharing**
- **For data consumers, ensure that data is served as desired**

Federated Data Product Lifecycle

1. Data onboarding
Persist the data product according to storage policies
Domain experts supply policies (e.g., privacy transformations)
2. Publishing
Register the data product in a federation-wide catalog
Support consumer-aware policies (e.g., # records)
3. Sharing
Include constraints (e.g., policies, transformations, time)
Sign contract and provide it to all contained parties
4. Consumption
Run compulsory operations (e.g., transformations)
Optimize consumption by moving data and/or processing
5. Discontinue
Remove data product from catalog (inform consumers)
Delete data product (and all copies) from all locations

Serverless Data Exchange

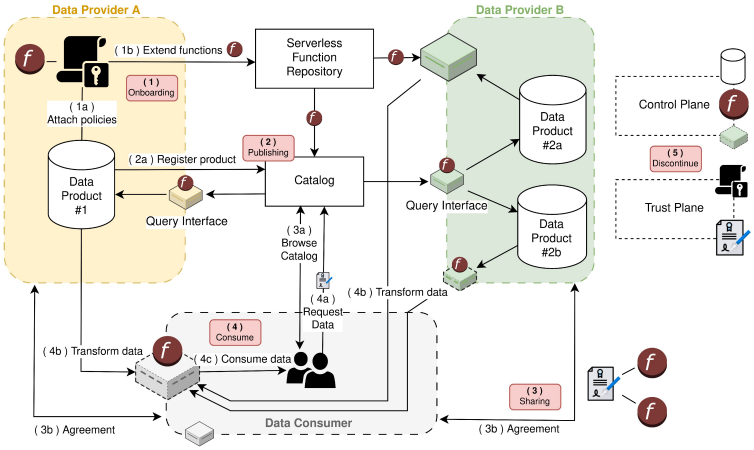

Fig. 2: Serverless data exchange of federated data products

- Supports the exchange of federated data products
- Follows the **federated data product lifecycle** (1-5)
- **Control plane** in charge of providing resources (i.e., processing, storage, serverless functions)
- **Trust plane** assures identities and policy enforcement
- Data is transformed ad hoc according to attached policies

Summary

- Data exchange between businesses hampered by data friction
- Providers and consumers require an architecture that supports **custom transformations** and ensures **privacy** policies
- Serverless functions (i.e., policies) are specified by partners and included as part of sharing agreements
- Transformations and identity are ensured automatically
- Thus, ensures privacy and simplifies data exchange workflow

Consortium

TEADAL

Funded by the European Union