FAIR Data

The data value chain of biotechnological processes

Alexandru Mardale
Advisor: Prof. Dr.-Ing. Dirk Weuster-Botz
Advisor: M. Sc. Lukas Bromig
Chair of Decentralized Systems Engineering
https://dse.in.tum.de/

15 November 2021 – 15 May 2022

Outline

- Motivation
- Background and related work
- Problem statement
- Design
- Implementation
- Evaluation

Motivation

- Large number of devices
- Large amounts of experiment data produced

Motivation

- Data reusability - context
  - Where did the data come from?
  - What experiment, device, device version?
  - When was it produced?

Outline

- ~~Motivation~~
- Background and related work
- Problem statement
- Design
- Implementation
- Evaluation

Background and related work

- FAIR Principles[1,2]
  - Findability
  - Accessibility
  - Interoperability
  - Reuse

[1]FAIR Principles: https://www.go-fair.org/fair-principles/

[2]Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Background and related work

- SiLA 2[1]
  - Common interface for variety of devices
  - Straightforward communication and operation of devices
- SiLA 2 Manager[2]
  - Central management for devices in laboratory
  - Handles device interaction
  - Handles experiment creation, scheduling and execution

[1]SiLA 2: https://sila-standard.com/standards/

[2]SiLA 2 Manager: https://gitlab.com/lukas.bromig/sila2_manager

Background and related work

- KNIME[1,2]
  - Modular analytics tools
  - Provides implementations for most common algorithms, techniques, input-output sources
  - Allows writing custom code
  - Community built modules
  - REST endpoints for analysis execution

[1]Michael R. Berthold, Nicolas Cebron, Fabian Dill, Thomas R. Gabriel, Tobias Kötter, Thorsten Meinl, Peter Ohl, Kilian Thiel, and Bernd Wiswedel. 2009. KNIME - the Konstanz information miner: version 2.0 and beyond. SIGKDD Explor. Newsl. 11, 1 (June 2009), 26–31. https://doi.org/10.1145/1656274.1656280

[2]KNIME: https://www.knime.com/

Outline

- ~~Motivation~~
- ~~Background and related work~~
- Problem statement
- Design
- Implementation
- Evaluation

Problem statement

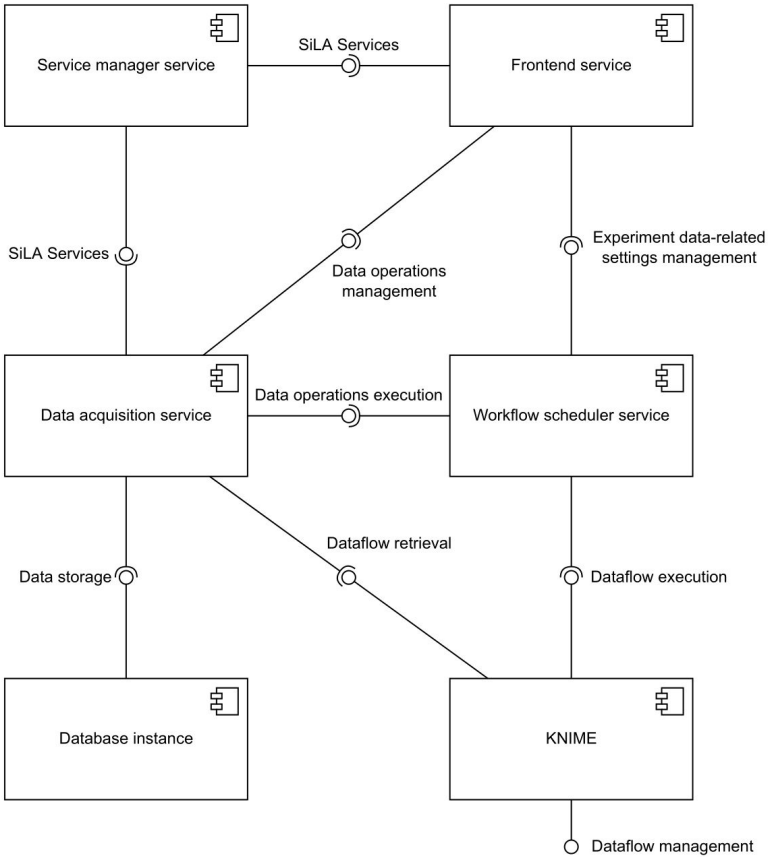
- Automate retrieval, storage and processing of experiment data
- Customize
  - what data is stored
  - how often data is stored
  - where data is stored
  - how data is processed
- Store metadata for reusability

Outline

- ~~Motivation~~
- ~~Background and related work~~
- ~~Problem statement~~
- Design
- Implementation
- Evaluation

Design

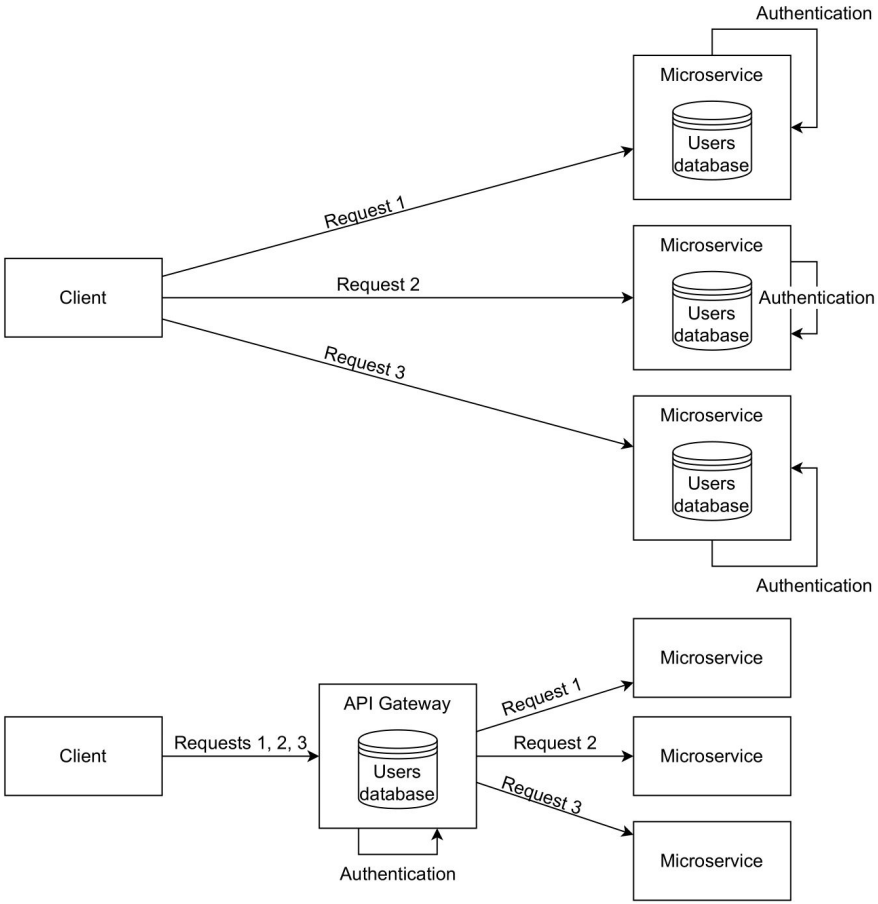
Microservice architecture



Design

API Gateway Pattern

- Route requests from clients to services
- Perform user authentication
  - Remove user duplication

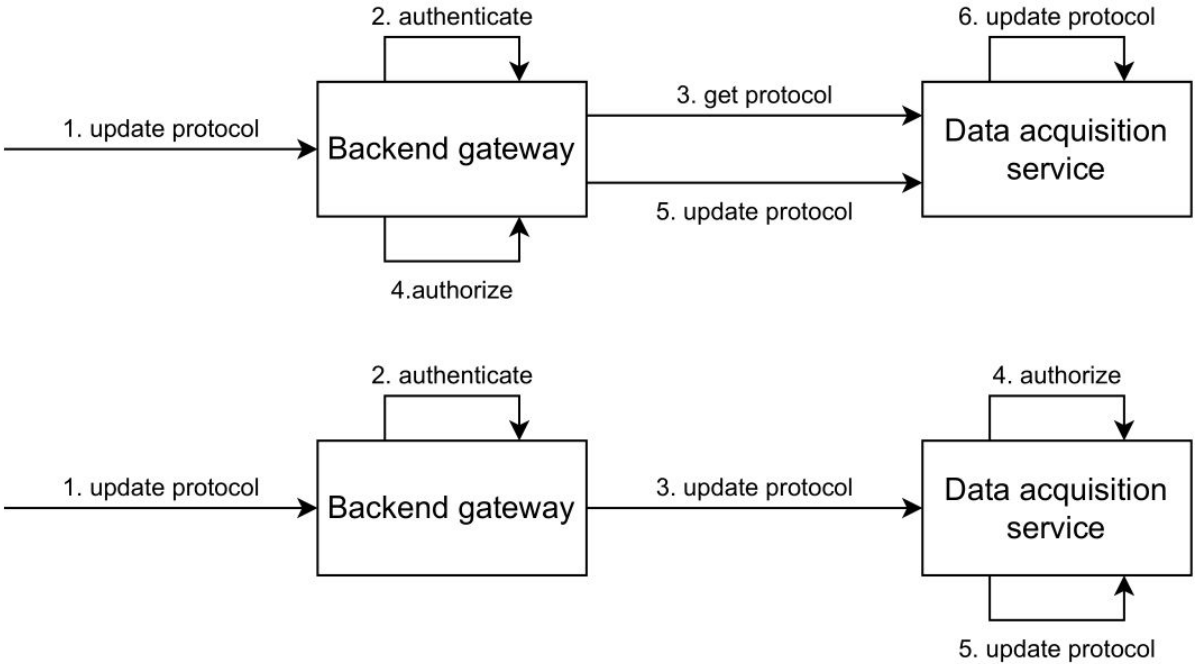


Outline

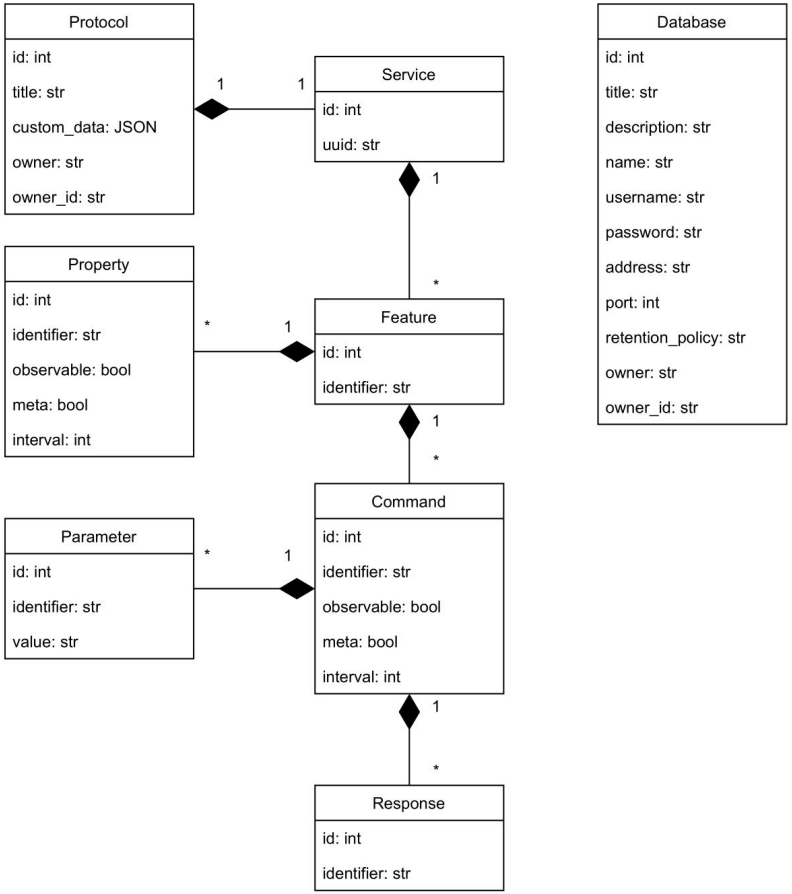
- ~~Motivation~~
- ~~Background and related work~~
- ~~Problem statement~~
- ~~Design~~
- Implementation
- Evaluation

Implementation

Authorization - gateway or microservice

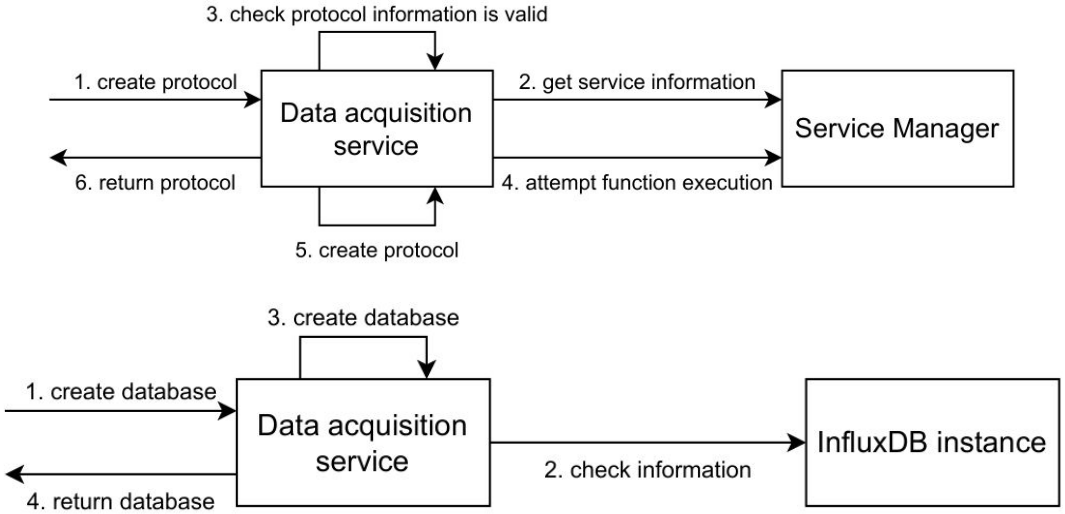


Implementation



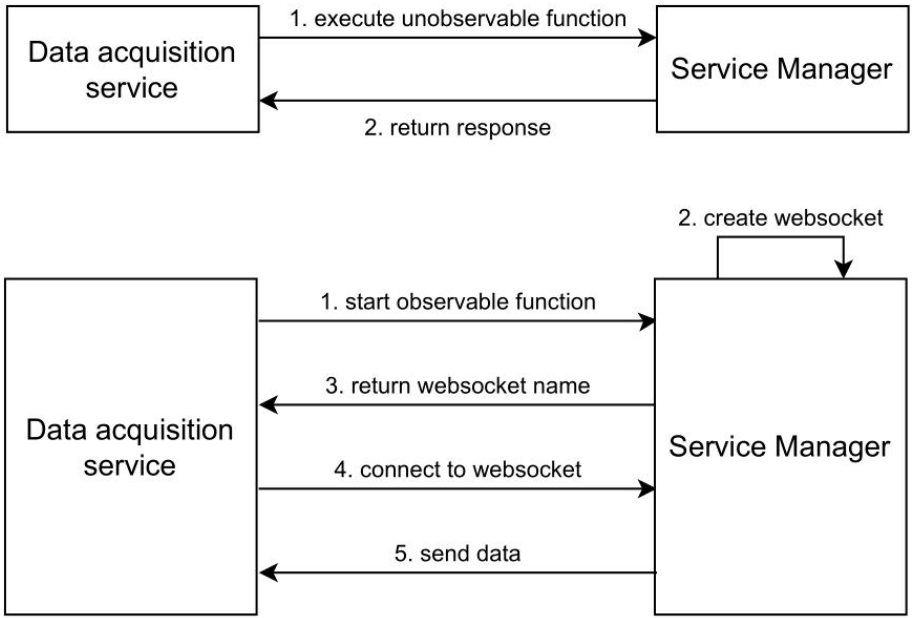
Implementation

Verifications performed on object creation / update



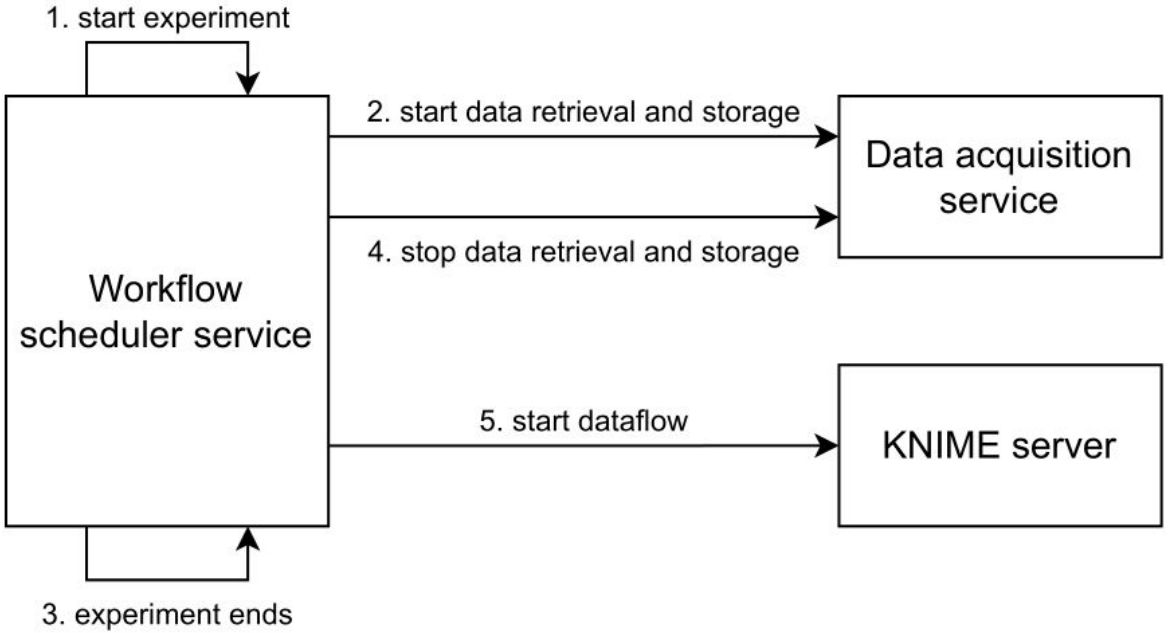
Implementation

Unobservable and observable function execution



Implementation

Data retrieval, storage and processing execution



Outline

- ~~Motivation~~
- ~~Background and related work~~
- ~~Problem statement~~
- ~~Design~~
- ~~Implementation~~
- Evaluation

Evaluation

- Execution in parallel of multiple
  - Experiments
  - SiLA services
  - Protocols with multiple
    - Functions
    - Intervals
    - Pieces of custom data
- Multiple databases
- Multiple dataflows

Thank you!