

# Data and Compute Federation at the Swiss Data Science Center



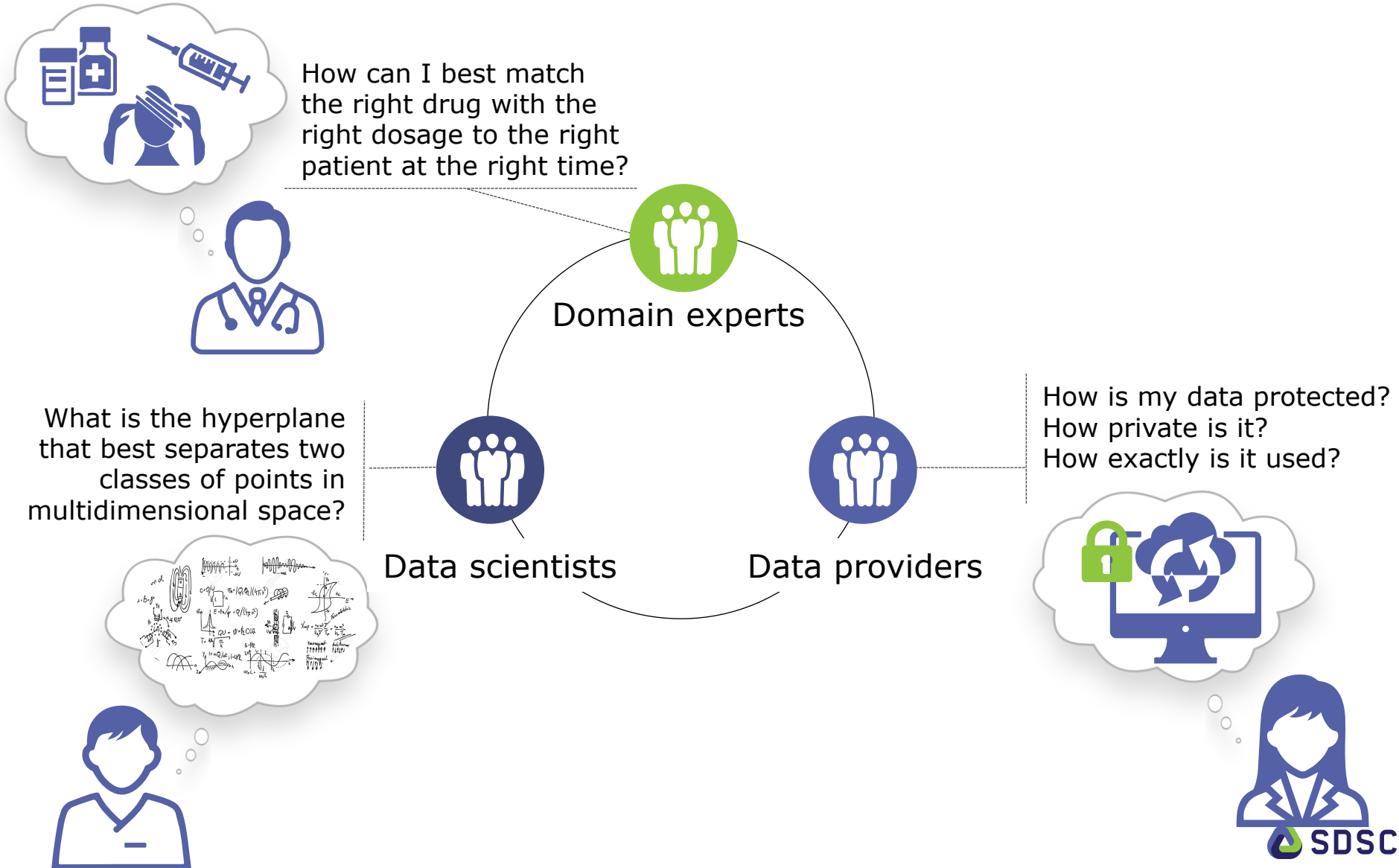
Sofiane Sarni, Ph.D.

Program Manager, SDSC

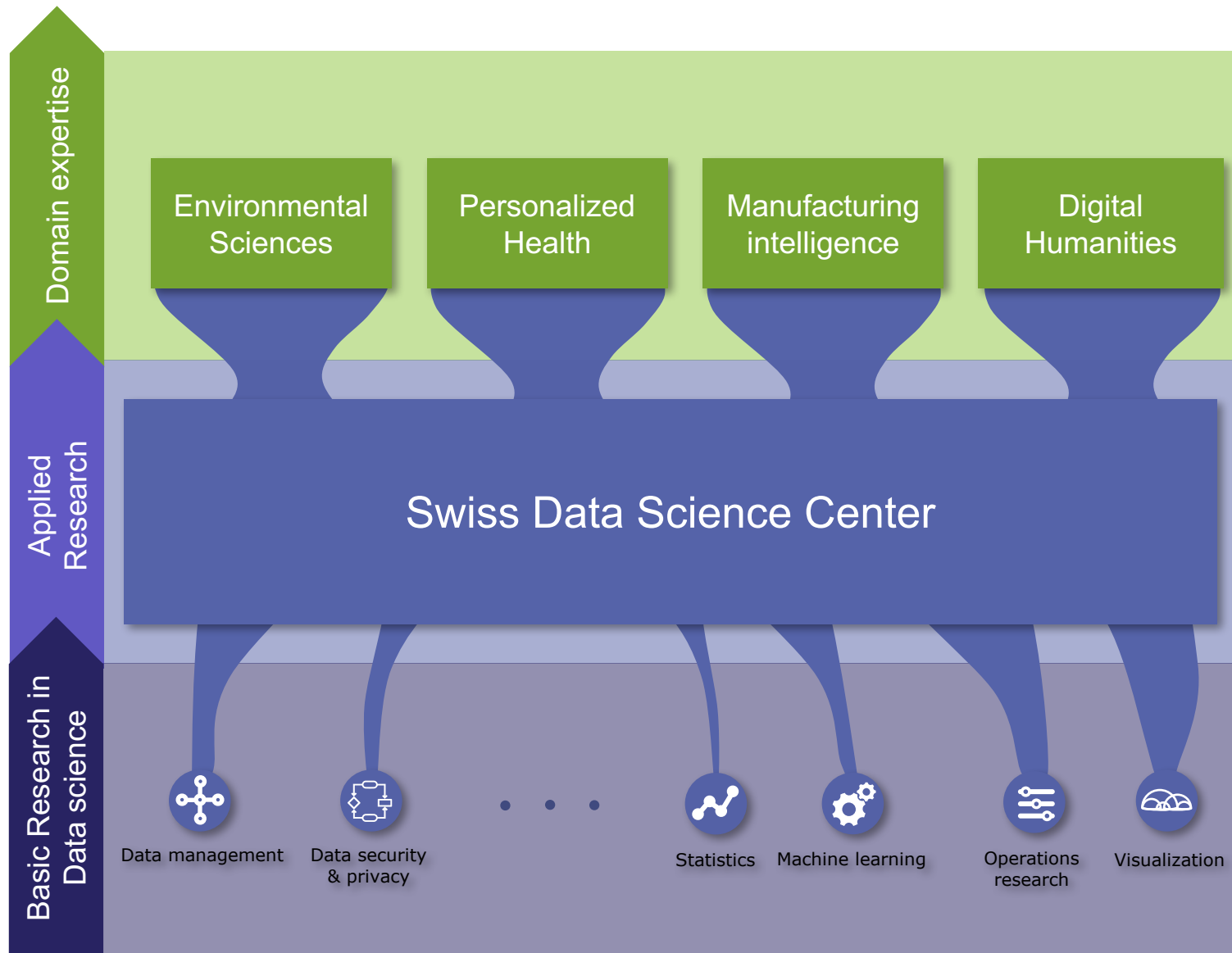
@SofianeSarni

# Swiss Data Science Center (SDSC)

Foster adoption of data science both in academia and industry

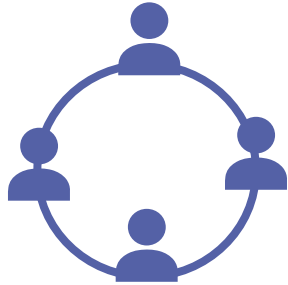


# Where does SDSC fit?



# What will the SDSC offer?

Excellence in academic research backed by strong industrial experience



## Embedded R&D collaboration

We engage in academic and industrial collaborations requiring large-scale distributed data processing (Big & Fast Data) and/or advanced analytics (machine learning & statistics) combined with an in-depth knowledge in select domains



## Domain-specific Insights as a Service

We provide secure access to our cloud-hosted analytics platform - the Renga platform (連歌), a highly scalable open software platform offering a one-stop-shop for hosting and exploring curated, calibrated and possibly anonymized data at scale, at-rest or in-motion.



## Open (Data) Science

The Renga (連歌) platform offers user-friendly tooling and services to help with the adoption of Open Science, fostering research productivity and excellence.

# Answering Researchers Challenges

- A data lake, not a data swamp!
    - Where can I upload my data, and make it available?
    - What other data is available? And where is it?
    - How was this data created? Who created it?
    - How trustful is it? Can I build my research on it?
- ... impedes **collaboration** between scientists, and **reusability** and **reproducibility** of research
- Data science made simple & trustable
    - Combining human expertise and machine intelligence
    - Making learning methods robust against uncertainties
    - Designing methods for interpretable machine learning

# Hosted Analytics Platform - Renga (連歌)

- Highly-scalable open software platform offering domain-specific insights as a service, featuring:
  - Data protection and digital rights management
  - Secure computing across (semi-)autonomous entities
  - Reusable research data and reproducible science
  - Agile data science via interactive IDE for rapid R&D
  - Domain-specific analytics SDK and frameworks

# Current Status

The center is fully operational as of January 2017

## Center set-up

January 17

- Hiring R&D staff
- Developing hosted platform
- **Collaborating across the Swiss academic community**
  - Personalized health
  - Environmental science
- **Engaging with industry**
  - Preventive maintenance

## Call for Academic Research Proposals

March 2017

### Motivations

- Foster and accelerate the adoption of data science across the ETH Domain
- Promote Open Science

### Research themes

- Data science meets domain science
- Data science methods for the real-world

## SDSC Industry Day

November 2017

### Objectives

- Showcase R&D activities of the center
- Offer a platform for industry to engage with SDSC

# The Software Platform - Renga (連歌)

Users, data scientists

Domain specific notebook and analytics

SDK

Environmental Science

SDK

Transport

SDK

Personalized Health

SDK

Social science

SDK

SDSC and partners


Common API

Micro services


- Security
- Privacy
- Federation
- Reproducible research
- Data protection

play akka lagom

Data Processing Applications



Open Big Data Platform Stack



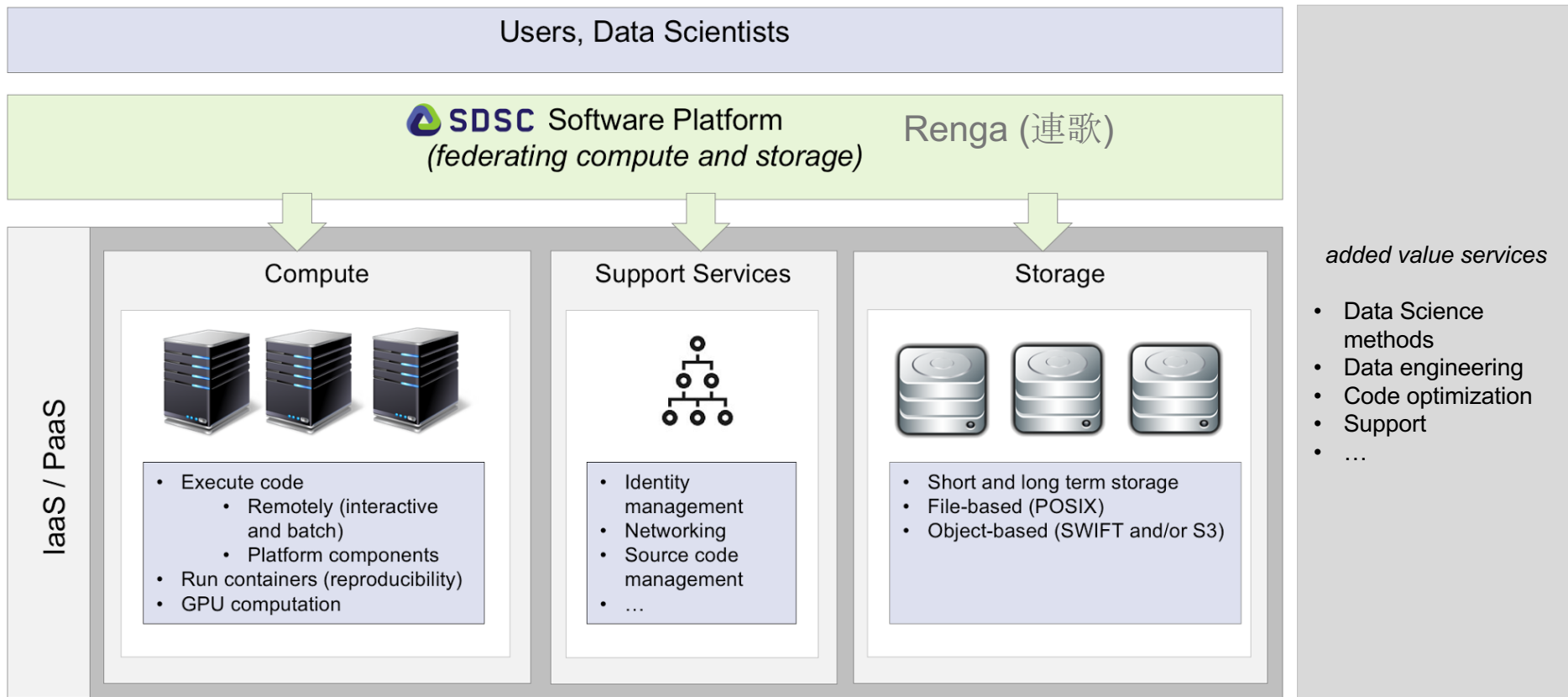
IaaS

Geographically distributed cloud and on premise infrastructure(s) + long term storage solution providers for archiving

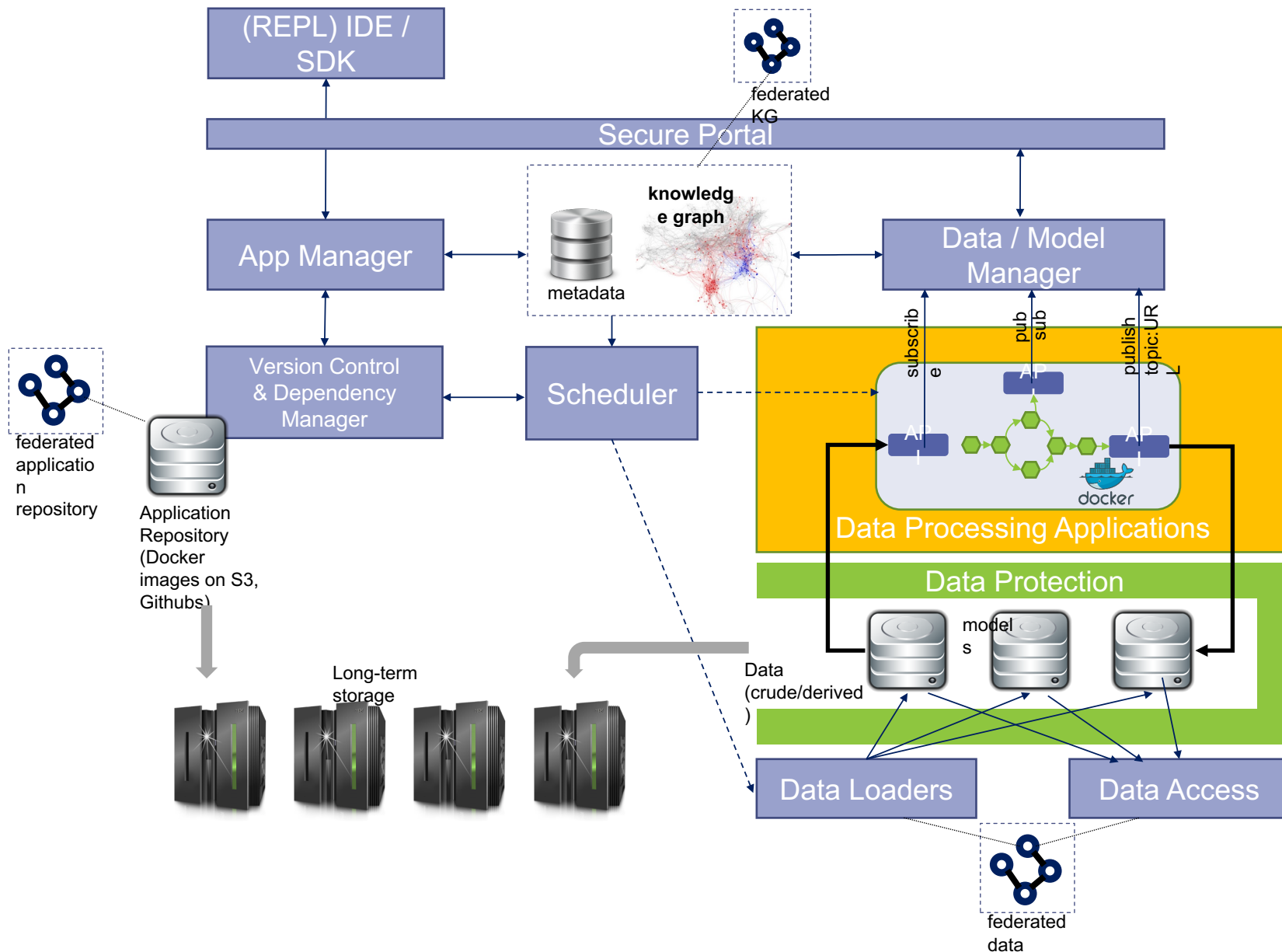




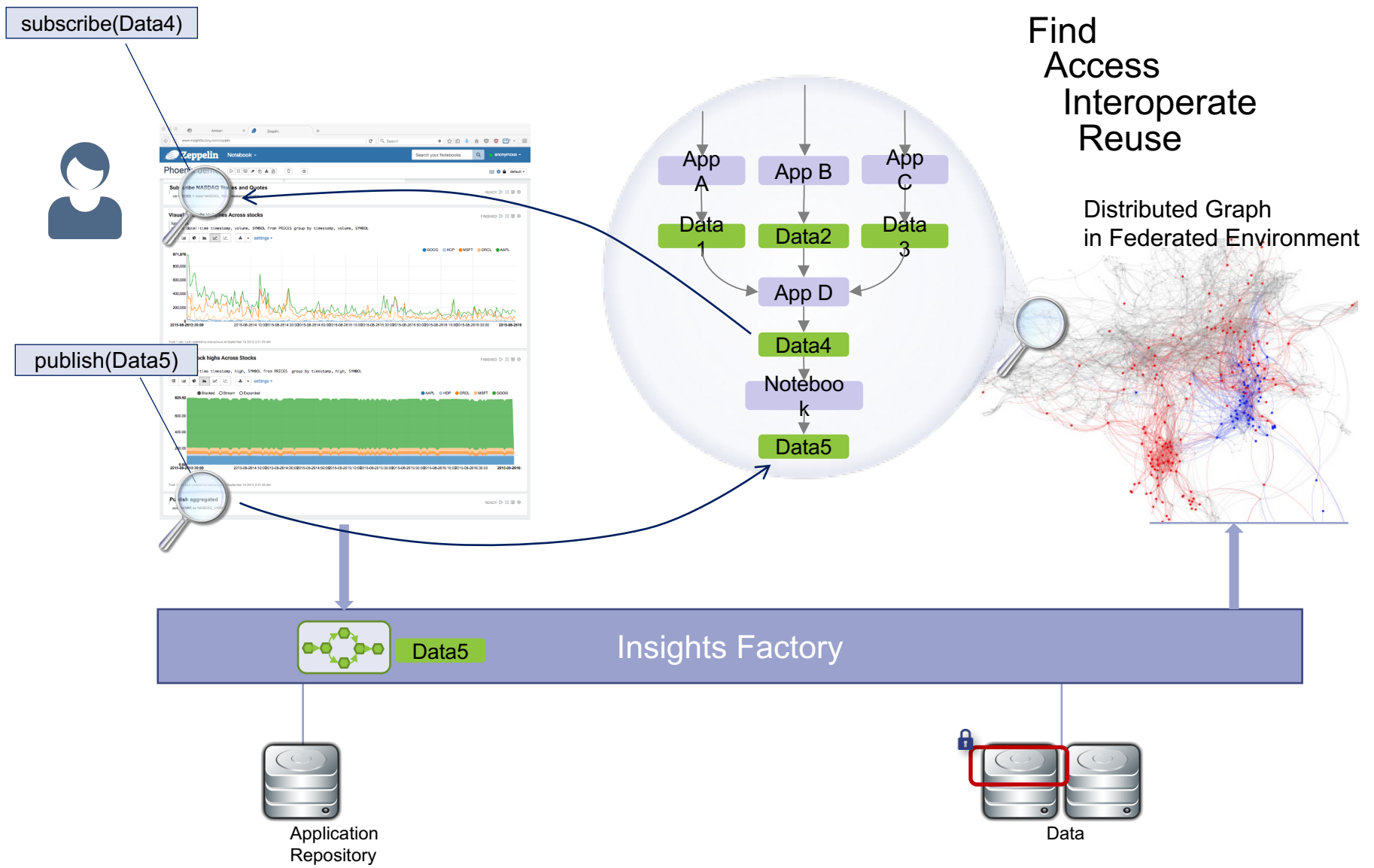
# Cloud Service Needs



# Building a Knowledge Graph

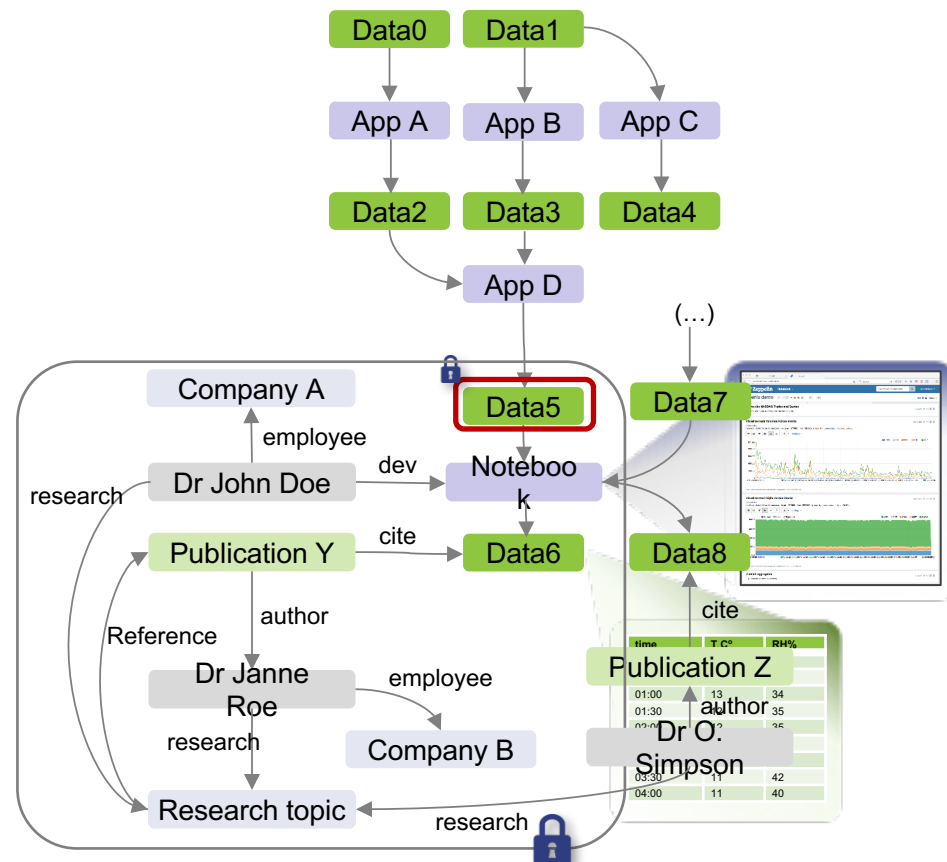


# Data Science Governance



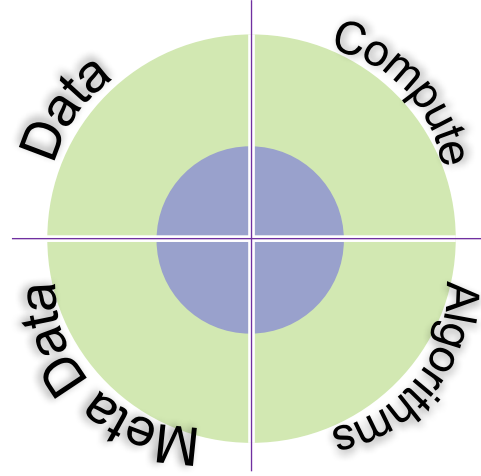
# Automated Open Science

- **Reproducible Research**
  - See the (versioned) algorithms
  - See the data
  - Replay a workflow
  - Compare workflows, validate robustness
- **Reusability**
  - Reuse data on new workflows
  - Clone and modify workflows
- **Knowledge Graph**
  - Data popularity, H-index
  - Who is using the data?
  - For what?
- **IP Protection**
  - Decide who sees the data,
  - The algorithms,
  - The data I use,
  - And how I use it

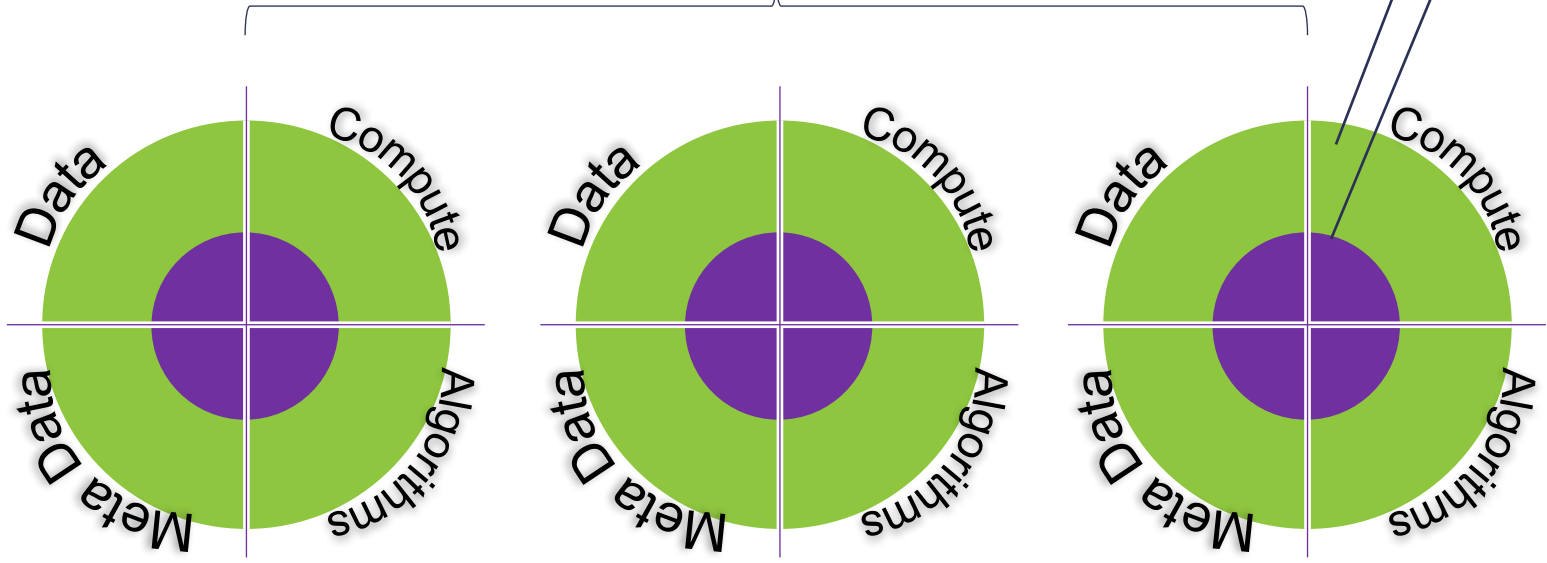


# Federation of Data and Compute

User perspective

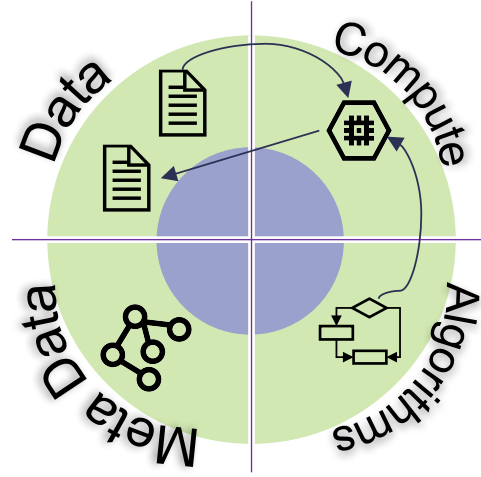


Federation perspective

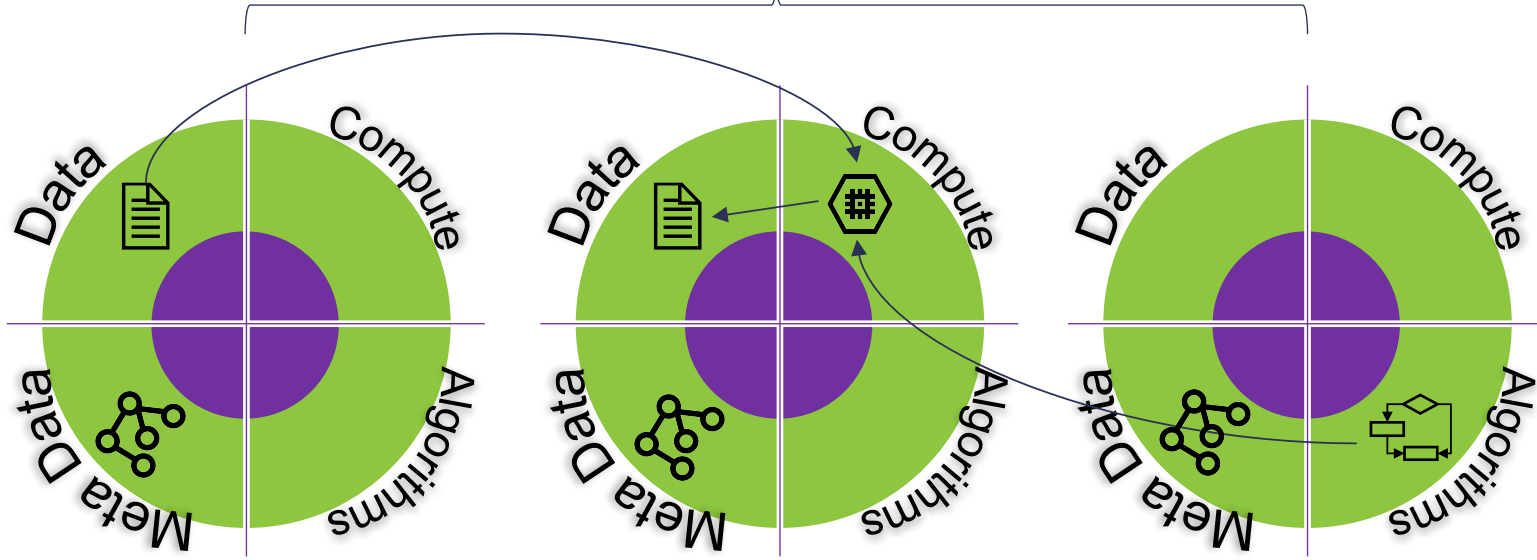


# Federation of Data and Compute

User perspective



Federation perspective



# Questions?



[datascience.ch](https://datascience.ch)



[sofiane.sarni@datascience.ch](mailto:sofiane.sarni@datascience.ch)



[@sofianesarni](https://twitter.com/sofianesarni)