Building a **Semantic** Layer for **Knowledge**-centric Research **Infrastructures**

Txell Amigó

Staff Scientist, Informatics Platform

Biotechnology Research Institute for the Green Transition (BRIGHT)



Option 1: Rose and sand, purple layer

Outline

1. WHY

Research Infrastructure Problems

2. WHAT

Our Solution: **Ontology**

2.1 Who we are: BRIGHT
2.2 Our solution: the PREFER ontology

3. HOW

Implementing it: **the Semantic Layer**

4. THE PREFER PROJECT

A **collaborative project**

WHY

Research Infrastructure **Problems**

Why am I here today?

Why ontologies?

Why semantic layers?

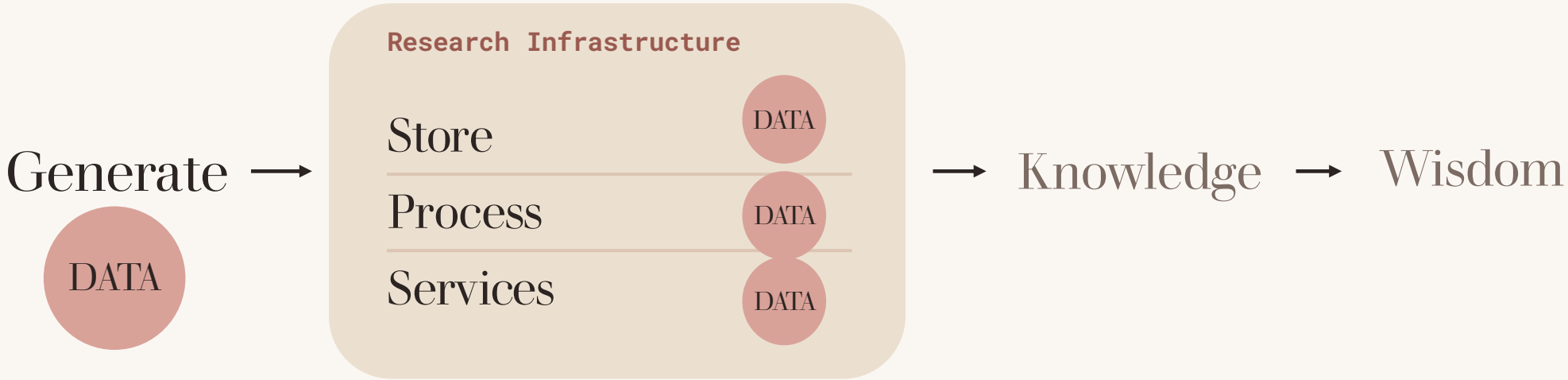
Why should you care?

Why does this matter for research infrastructures, and so for DeiC?

Diagnosing a digital infrastructure

The **ROLE** of a Digital Research Infrastructure:

manage
digital information → DATA



Diagnosing a digital infrastructure

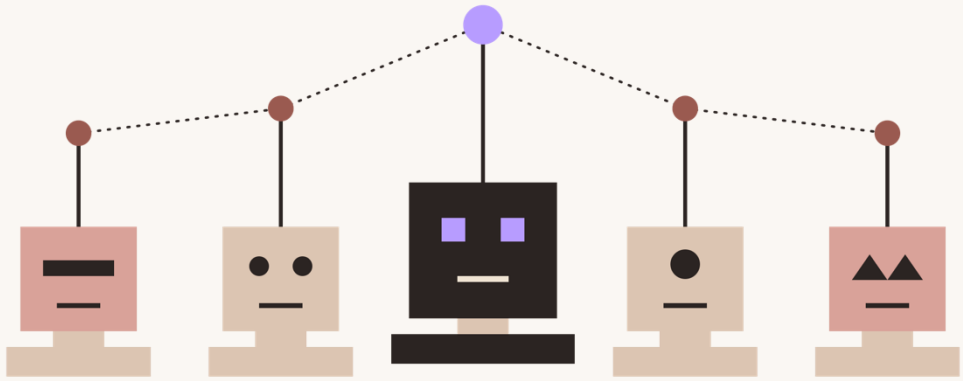
	Need	Task / PROBLEM
Legacy Data	Find a past experiment	Ask around, search folders (up to some hours)
	Answer data quality-control questions	Reconstruct from Electronic Lab Notebooks, no lineage or provenance
Collabs	Exchange data with a partner	Reformat and explain by hand, write another custom script
AI	Train a machine learning model	Harmonise a big pool of datasets (weeks of manual curation)

Diagnosing a digital infrastructure

AI

Need	Task / PROBLEM
Integrate agentic systems	Hallucinations: many tokens spent, no verification, no reliability

AI amplifies the problem



Diagnosing a digital infrastructure

Main problem:

Data integration problem

↓

Data fragmentation

↓

Data silos

Diagnosing a digital infrastructure

PROBLEM: Data silos

CAUSE: **LACK OF STANDARDS**

We all know the **mantra**: the **FAIR** Principles

Without shared standards, data cannot be made

F — Findable

A — Accessible

I — Interoperable

R — Reusable

But How?

Diagnosing a digital infrastructure

PROBLEM: ~~Data silos~~

what we see

Semantic silos

what is underneath

A **meaning** problem, a **context** problem,
and all the **logic** that comes with the data.

To bridge semantic silos → **Ontologies**

Diagnosing a digital infrastructure

	Need	Task / PROBLEM
Legacy Data	Find a past experiment	Ask around, search folders (up to some hours)
	Answer data quality-control questions	Reconstruct from Electronic Lab Notebooks, no lineage or provenance
Collabs	Exchange data with a partner	Reformat and explain by hand, write another custom script
AI	Train a machine learning model	Harmonise a big pool of datasets (weeks of manual curation)
	Integrate agentic systems	Hallucinations: many tokens spent, no verification, no reliability

Diagnosing a digital infrastructure

	Need	Task / PROBLEM	**Task with an ontology**
Legacy Data	Find a past experiment	~~Ask around, search folders (up to some hours)~~	**One query across all storage systems**
Legacy Data	Answer data quality-control questions	~~Reconstruct from Electronic Lab Notebooks, no lineage or provenance~~	**Follow the links from result to source**
Collabs	Exchange data with a partner	~~Reformat and explain by hand, write another custom script~~	**Export in standard, documented terms**
AI	Train a machine learning model	~~Harmonise a big pool of datasets (weeks of manual curation)~~	**Pull a consistent, labelled dataset**
AI	Integrate agentic systems	~~Hallucinations: many tokens spent, no verification, no reliability~~	**A more resilient, less expensive system; explainable outputs linked to data sources**

WHAT

Is Our Solution: **Ontology**

2.1 Who we are: BRIGHT

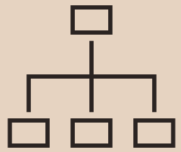
2.2 What is the PREFER ontology

What is an ontology?



Dictionary

Terms and their descriptions.



Taxonomy

Plus synonyms and a hierarchy of terms.



Ontology

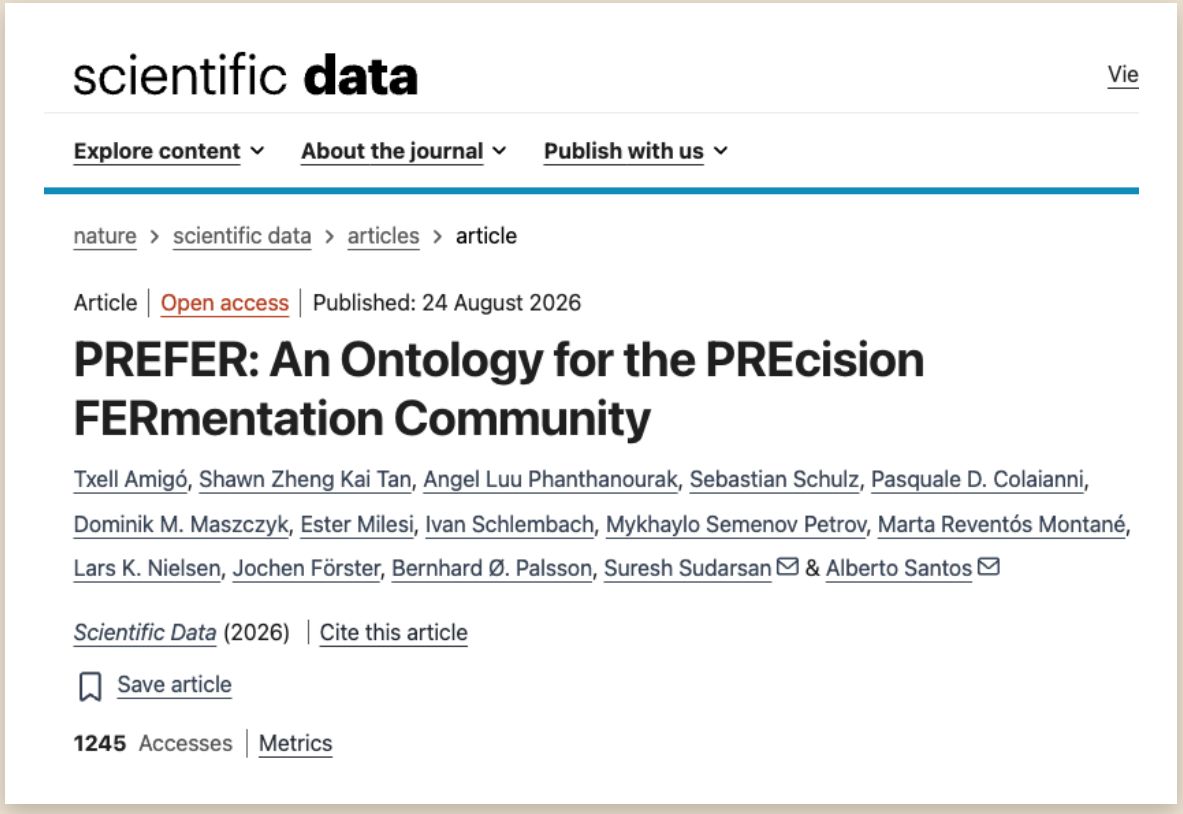
Plus domain-specific relationships between terms.

An ontology is a structured framework that defines the concepts, properties and relationships within a domain.

PREFER Ontology

An ontology for the **PR**Ecision **FER**mentation community

Open-source ontology designed to establish a unified standard for bioprocess data, covering the whole precision fermentation process.



scientific **data**

Vie

Explore content | About the journal | Publish with us

nature › scientific data › articles › article

Article | Open access | Published: 24 August 2026

PREFER: An Ontology for the PREcision FERmentation Community

Txell Amigó, Shawn Zheng Kai Tan, Angel Luu Phanthanourak, Sebastian Schulz, Pasquale D. Colaianni, Dominik M. Maszczyk, Ester Milesi, Ivan Schlembach, Mykhaylo Semenov Petrov, Marta Reventós Montané, Lars K. Nielsen, Jochen Förster, Bernhard Ø. Palsson, Suresh Sudarsan & Alberto Santos

Scientific Data (2026) | Cite this article

Save article

1245 Accesses | Metrics

Published in Scientific Data (Nature), 2026

Who we are: BRIGHT

Novo Nordisk Foundation Biotechnology Research Institute for the Green Transition



Core Groundbreaking biotech research and cutting-edge technology.

Aim Scalable solutions, brought to market with industry, as competitive alternatives to fossil-based products and processes.

Three main missions:



Sustainable Materials



Microbial Foods



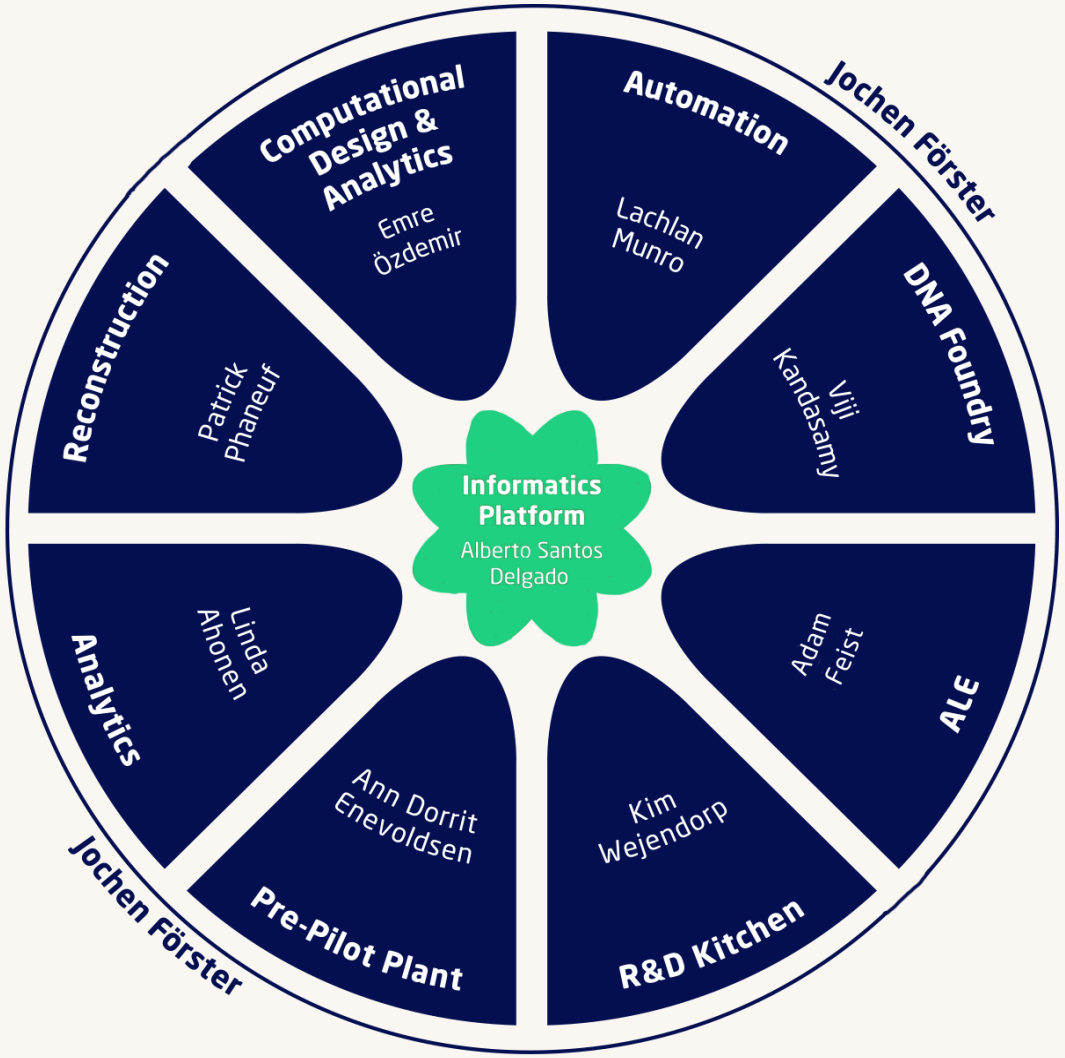
Net Zero Agriculture

BRIGHT Biofoundry

Led by Jochen Forster

A unified **infrastructure of biotechnology laboratories** that integrates robot-based workflows and technology platforms to **accelerate innovation** in **biocatalyst-based production.**

Biological factories for biomanufacturing: whole cells or subcellular fractions.



The Informatics Platform at BRIGHT



Standards Office


Lea M. Sommer


Mia Lopez


Ding He


Txell Amigó

Data Science Platform


Albert Pallejà


Juliana Assis


Sebastian Schulz


Henry Webel

Research Software Engineering


Pasquale Colaianni


Fatemeh Ganami


Vadim Tofan


Atieh Gharib


Dominik Maszczyk

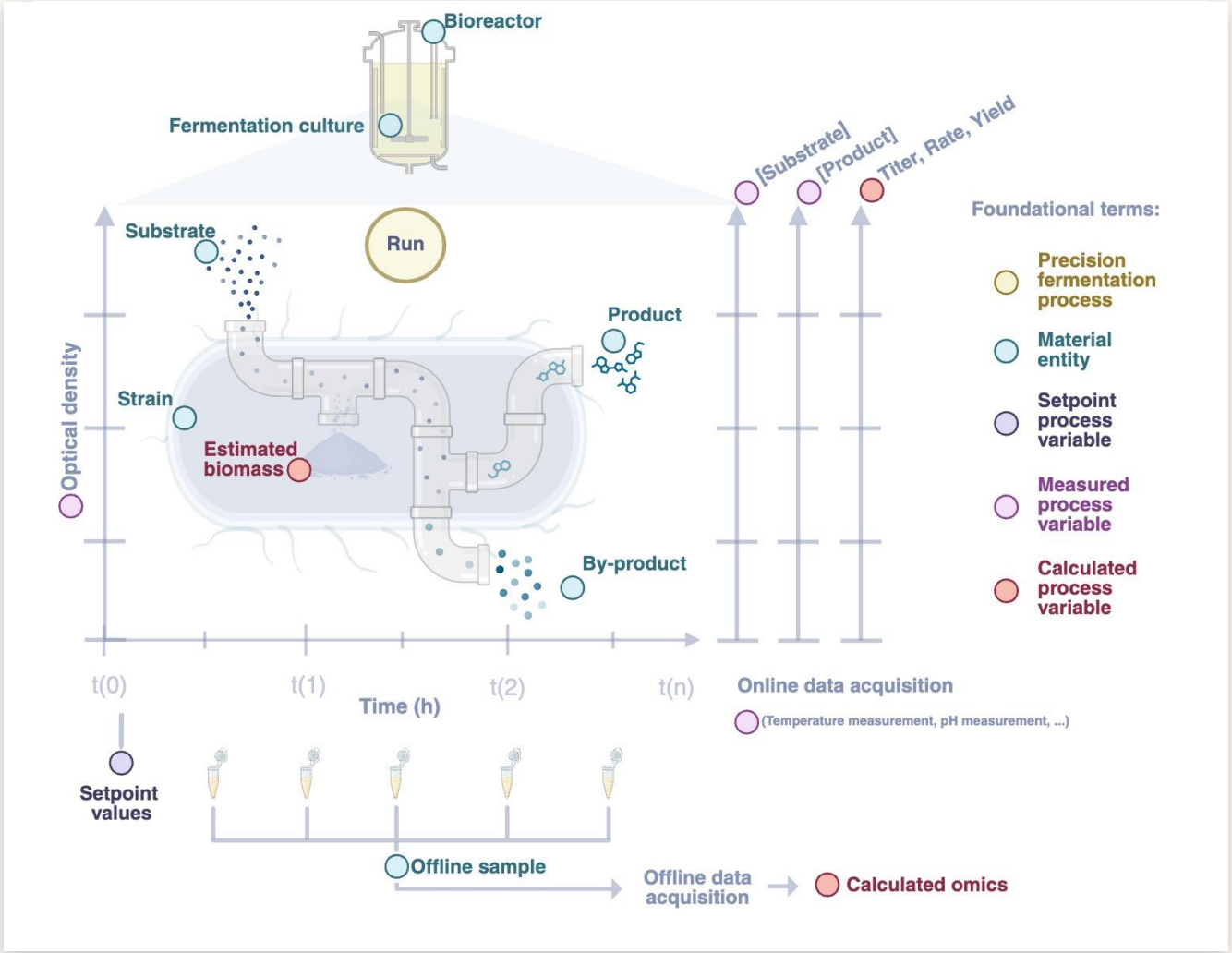

Vasileia Vagena



Alberto Santos
Platform lead

PREFER Ontology

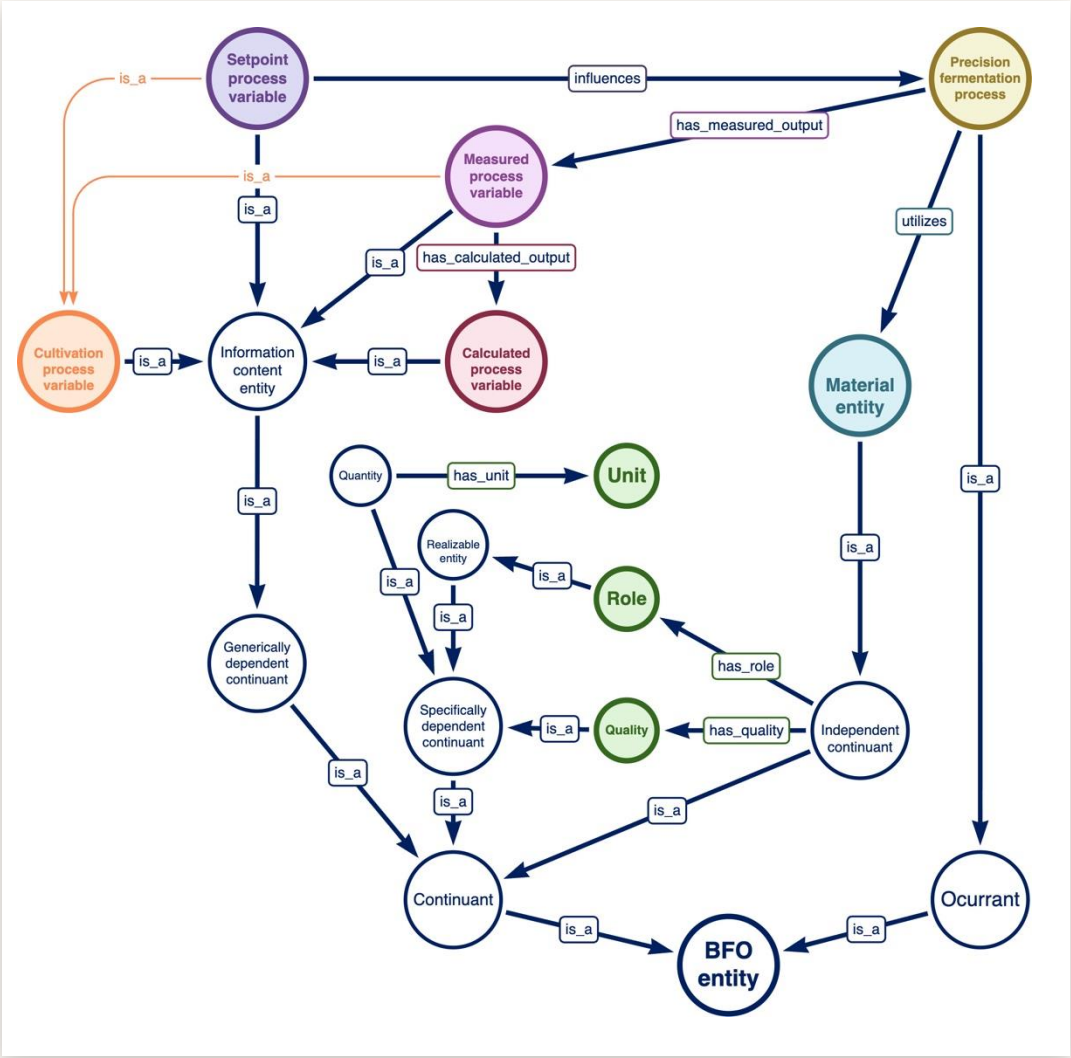
Figure 1. **A journey of a single bioreactor run**: exploring a precision fermentation process.



The ontology represents **the whole run process**: setpoints, measurements, calculated variables.

PREFER Ontology

Figure 2. **Overview of the foundational terms of the PREFER ontology** and their alignment with the Basic Formal Ontology (BFO).



Building on BFO, an upper-level ontology, makes PREFER interoperable with other ontologies.

PREFER Ontology

A standard, built in a standardised way

odk
ONTOLOGY DEVELOPMENT KIT

Open-source toolbox and ontology life-cycle management system for a standardised, transparent development process.

Reusability / Interoperability / Best practices / Automated QC



OBO Foundry

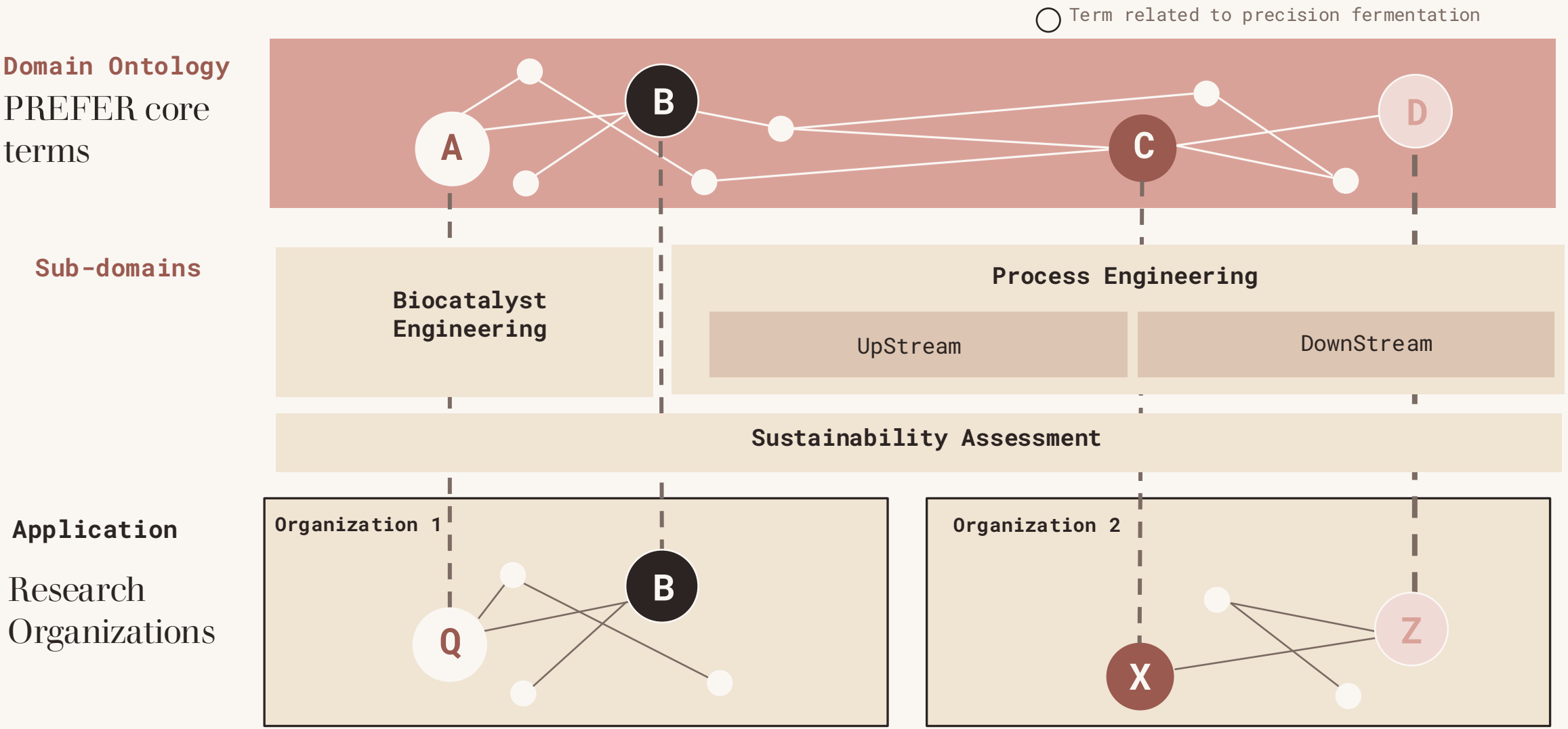
Community development of interoperable ontologies, with quality control and unique, non-overlapping scopes.

PREFER is an accepted ontology.

Open Biological and Biomedical Ontology / obofoundry.org

DeiC

PREFER Ontology: **Scope**

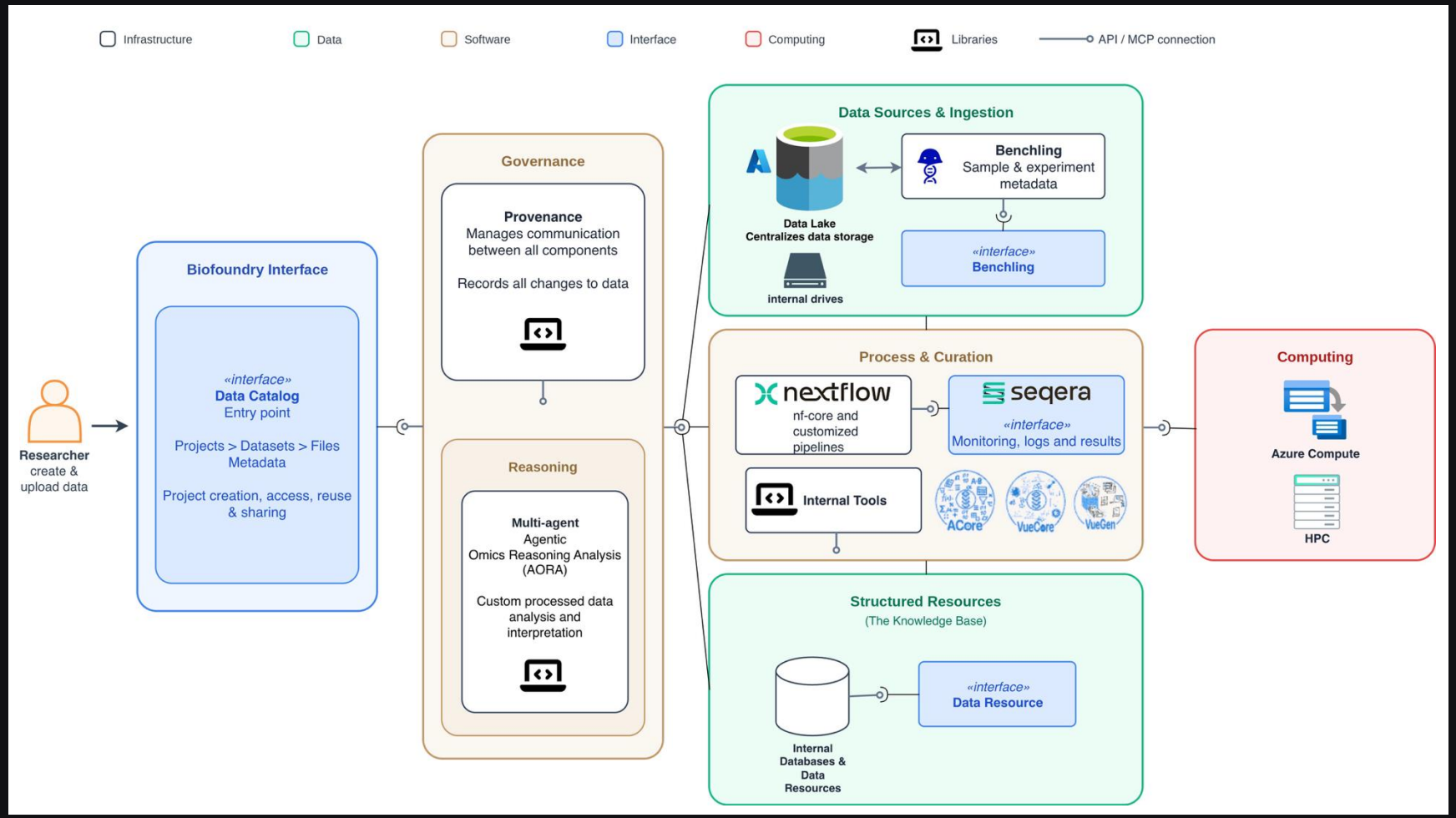


HOW

How semantic data flows through the digital infrastructure?

Implementing it: **the Semantic Layer**

BRIGHT Data Infrastructure



Biofoundry Interface
entry point for researchers

Governance
provenance, and Reasoning Box (agentic data analysis)

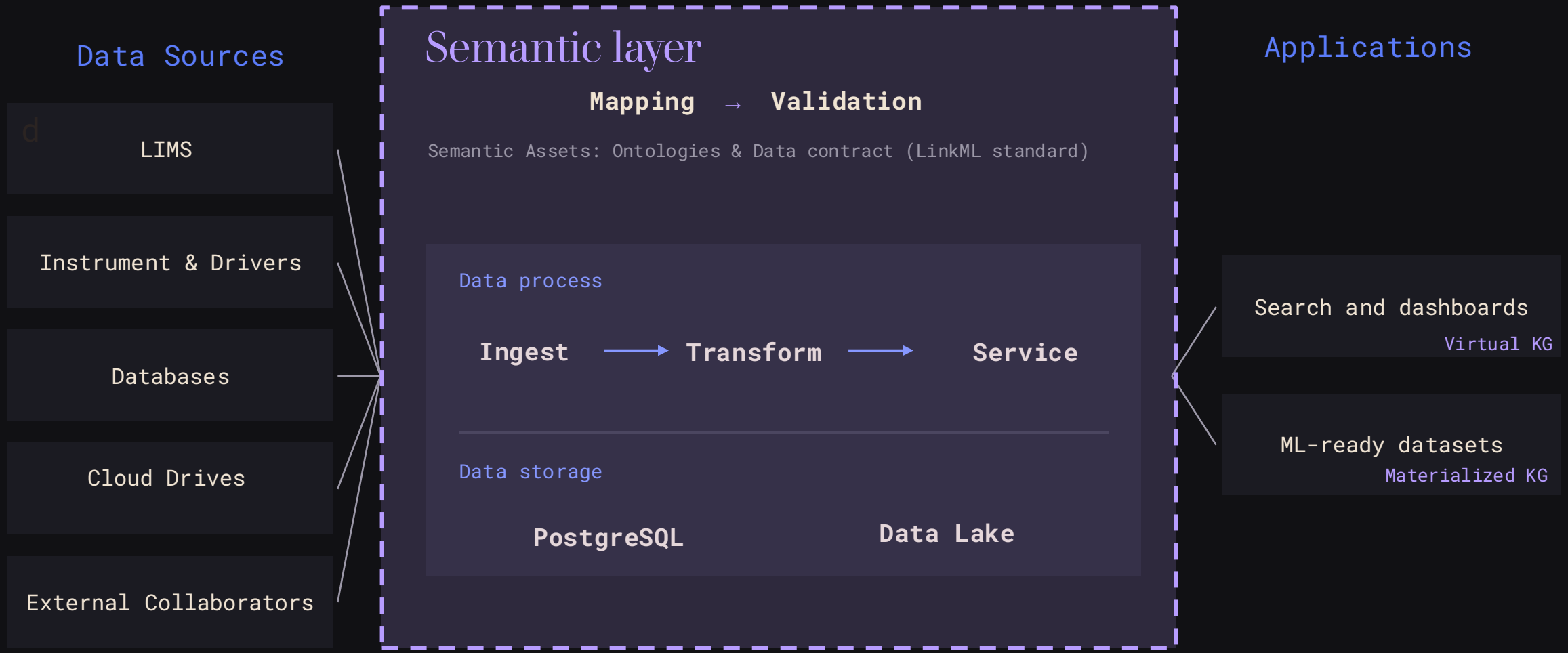
Data Sources & Ingestion
Data and metadata

Process & Curation
Pipelines from Nextflow and nf-core on Azure

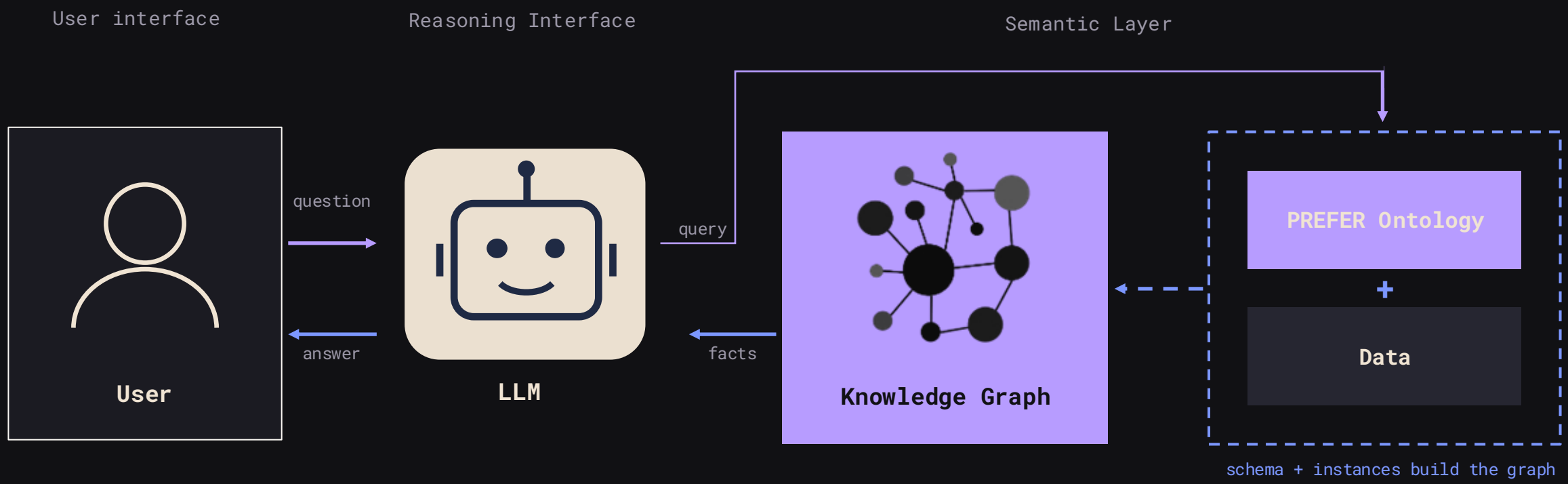
Computing
Azure Compute & HPC (DTU)

Structured Resources
internal databases

The semantic layer, summarised



Use case



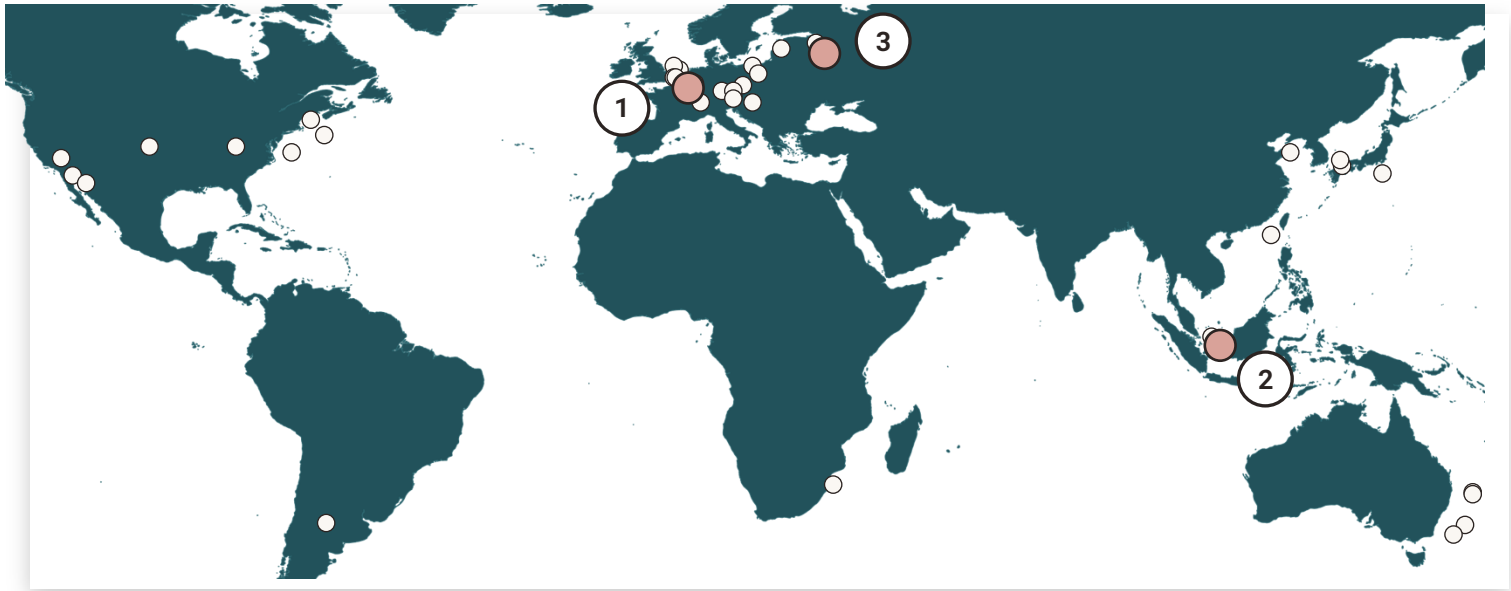
Translation layer Accurate querying, fewer hallucinations: the LLM is constrained to known facts.

The PREFER Project

A **collaborative** project

The PREFER Project

41 members of the **Global Biofoundry Alliance** across 17 countries



PREFER collaborator



Lea M. Sommer
PREFER Project
Collaboration Leads

Global Biofoundries Alliance

1 London, United Kingdom

LONDON BIO FOUNDRY
Imperial College London

2 Singapore, Singapore

BioFoundry SINGAPORE
NUS National University of Singapore
SynCTI Synthetic Biology for Clinical & Technological Innovation

3 Tartu & Tallinn, Estonia

DIGIBIO
TAL TECH Tallinn University of Technology
University of Tartu

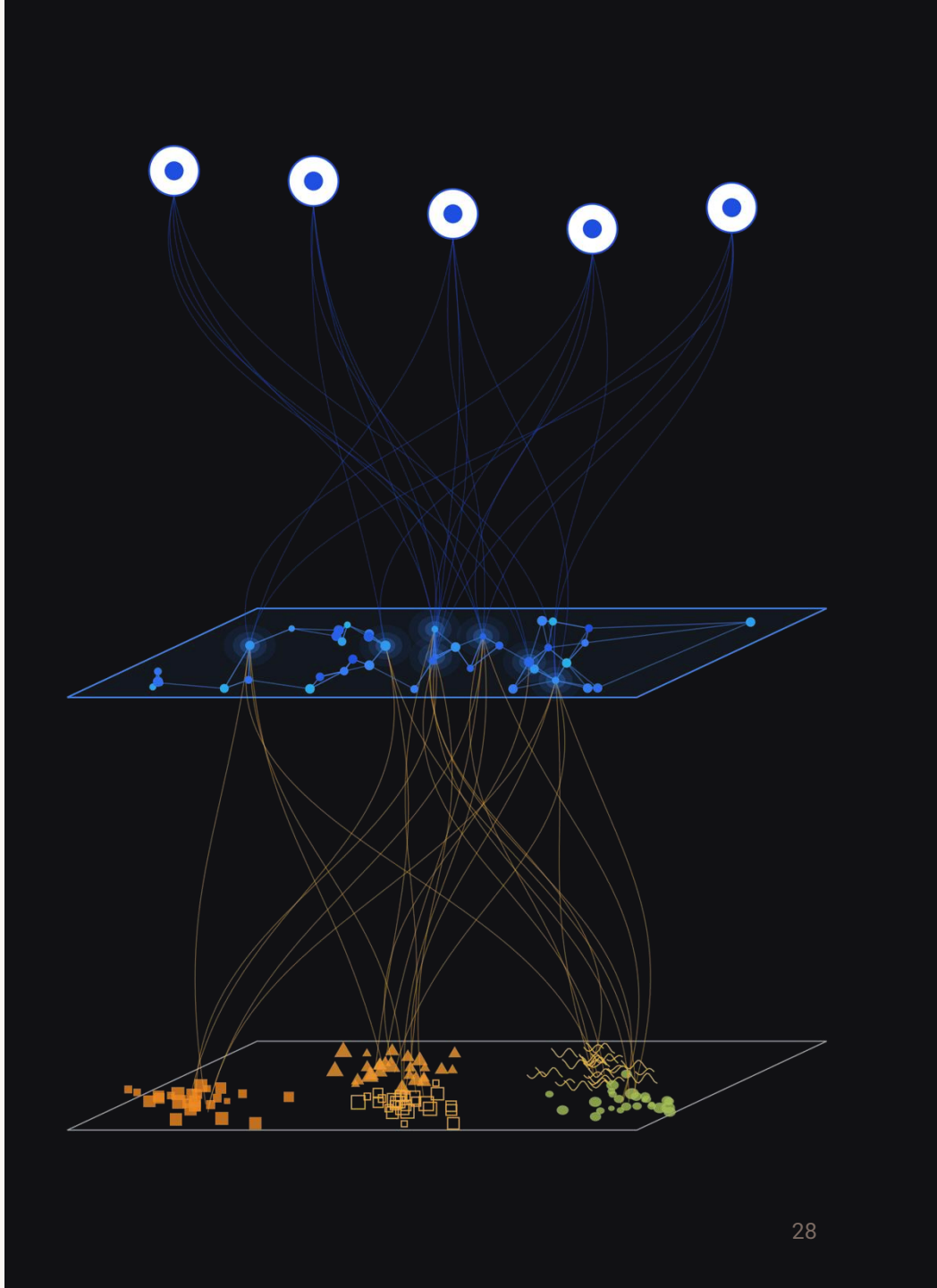
Source: biofoundries.org/members-1 (Oct 2026). Marker positions from host institution locations; RISE and Núcleo de Innovación de Sistemas Biológicos placed approximately.

DeiC

DeiC

“Data quality is a consequence of collaboration quality”

The hidden value of an ontology is the community



If you want to know more...

GitHub repository



github.com/Multiomics-Analytics-Group/prefer_ontology

Manuscript



nature.com/articles/s41597-026-08099-x

Ontology Web Browser



ebi.ac.uk/ols4/ontologies/prefer

Thank you



BRIGHT
BIOFOUNDRY














novo nordisk

fonden

Benefitting people and society