



**Ontologies in Pharma:
The landscape, pre-competitive
development, and use-cases**

Becky Upton
Pistoia Alliance
President



Ontologies in Pharma: The landscape, pre-competitive development, and use-cases

Dr Becky Upton



A data explosion



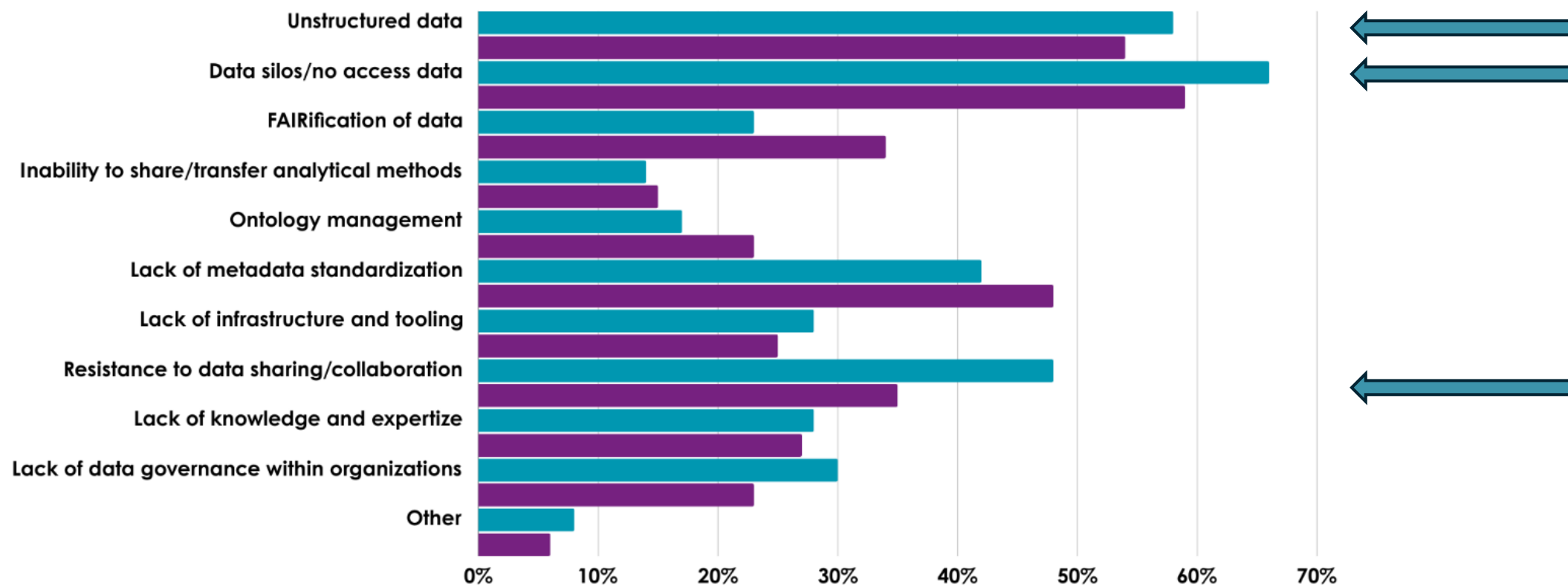
Data is the new gold?

“Digital transformation is fueled by data, but data do not proliferate neatly. The more of it that gets produced, the more challenges emerge around infrastructure, processing, cleaning, wrangling, and sharing”

Pistoia Alliance LOTF survey



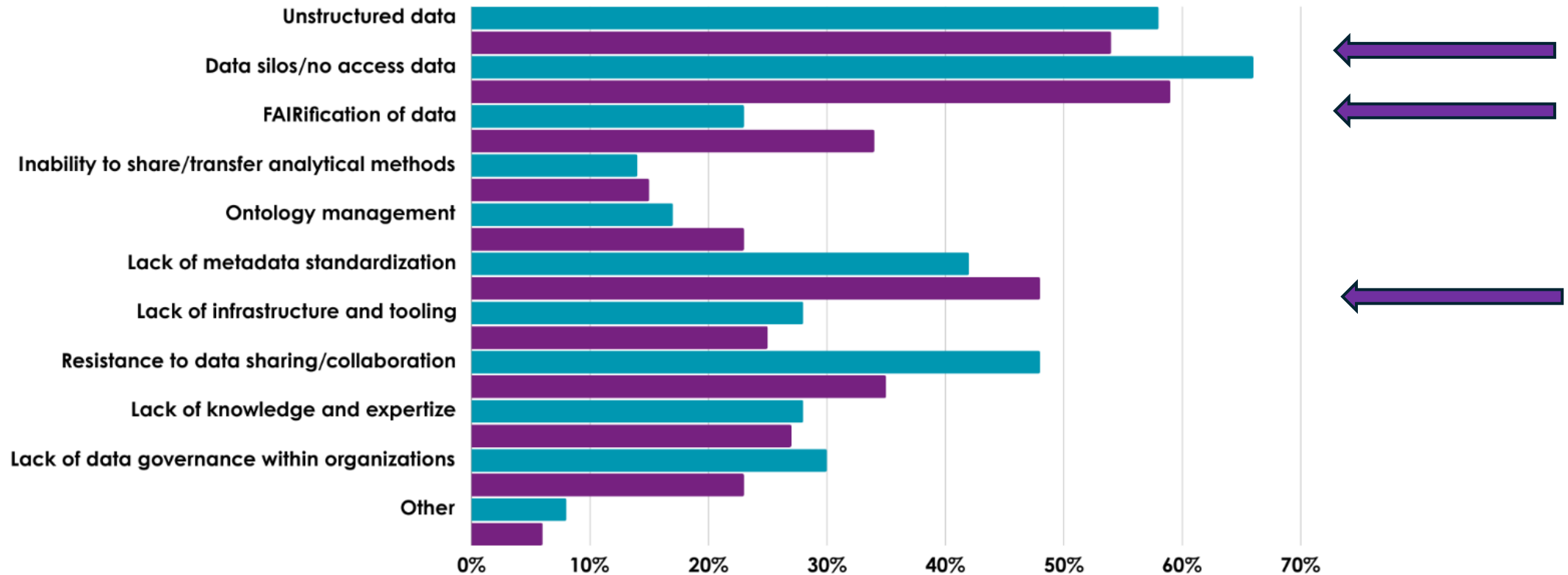
What are the biggest barriers to making the best use of experimental data? (Tick your top three)



Pistoia Alliance LOTF survey



What are the biggest barriers to making the best use of experimental data? (Tick your top three)



FAIR data

F A I R



Findable



Accessible



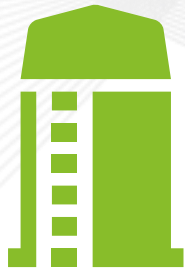
Interoperable



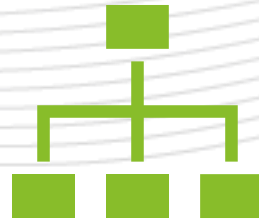
Reusable

The FAIR Guiding Principles for Scientific data management and stewardship – Wilkinson, Dumontier, Moens etc al.
Nature, 2016

Data silos



Traditional data silos – hard to get data out, no standards applied

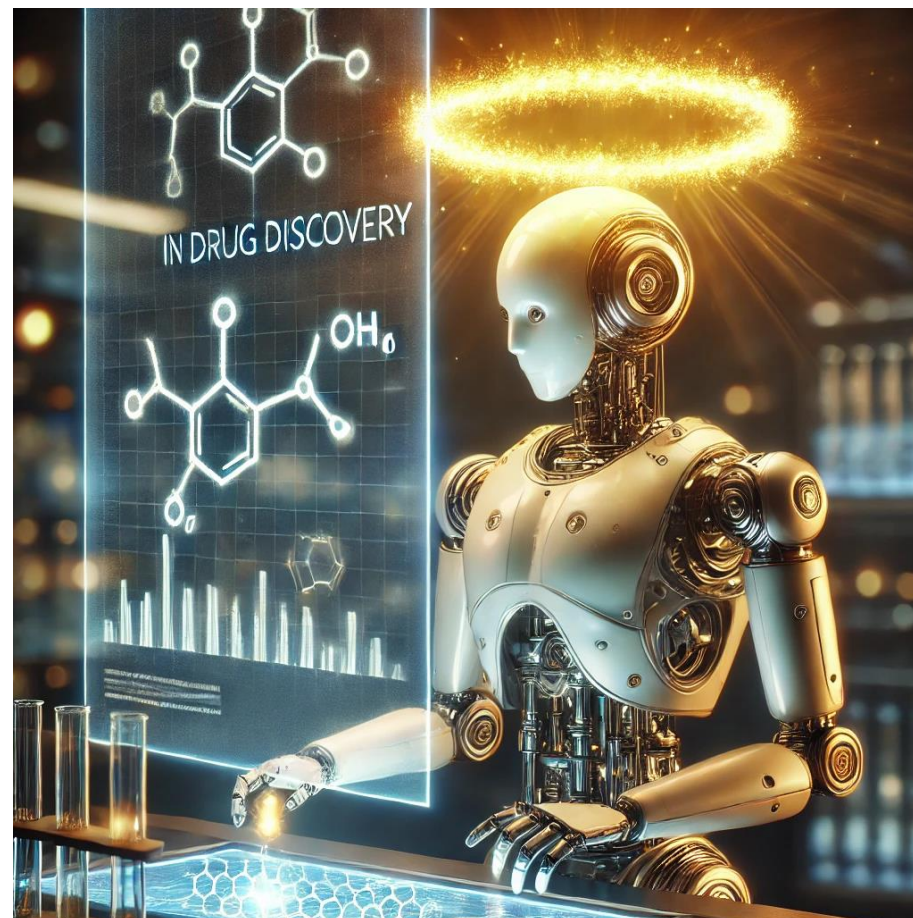


'FAIR data silos' – standards applied but only within that platform



Software platforms need to be able to ingest ontologies as well as data

AI – a saviour?



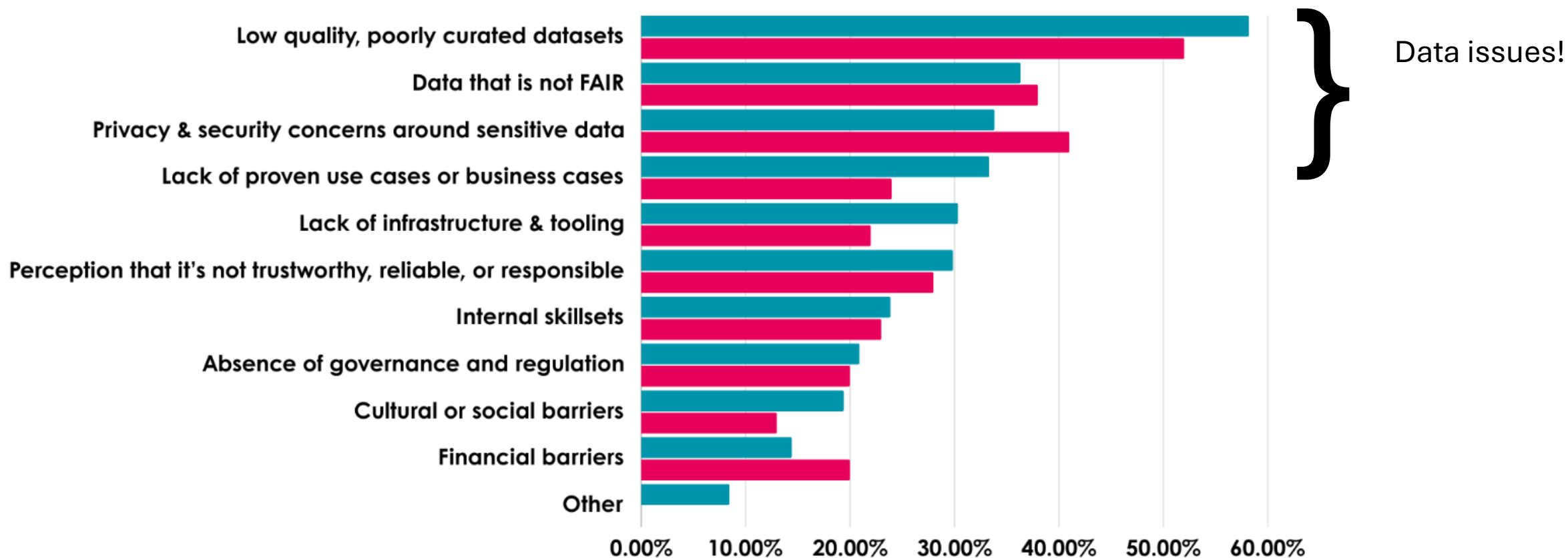
AI – a saviour? Not without issues

- Accuracy and reliability
- Interpretation of complex data
- Lack of specialized knowledge
- Bias in data
- Over confidence in results

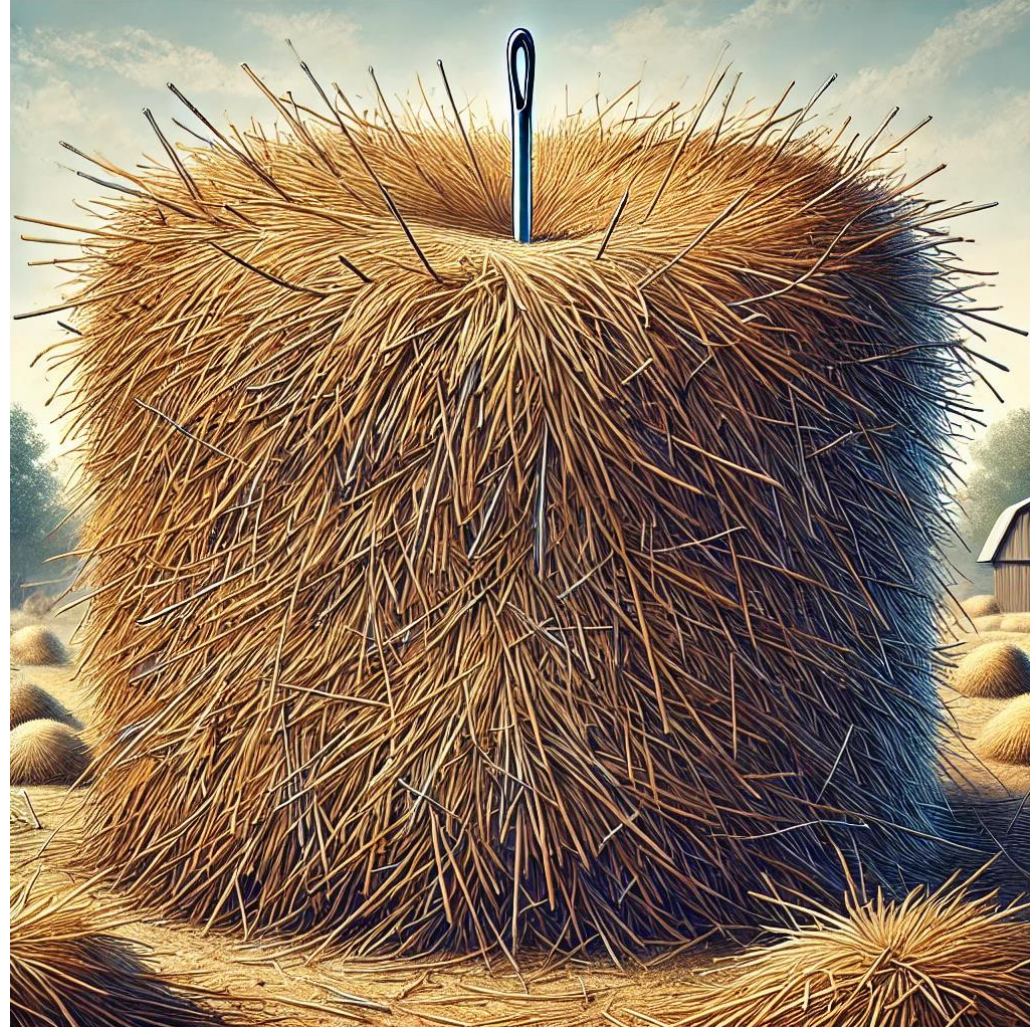
Pistoia Alliance survey



What are the biggest barriers to implementing AI/ML at scale within the laboratory environment? (Tick your top three)



How do we help LLMs and manage our data?

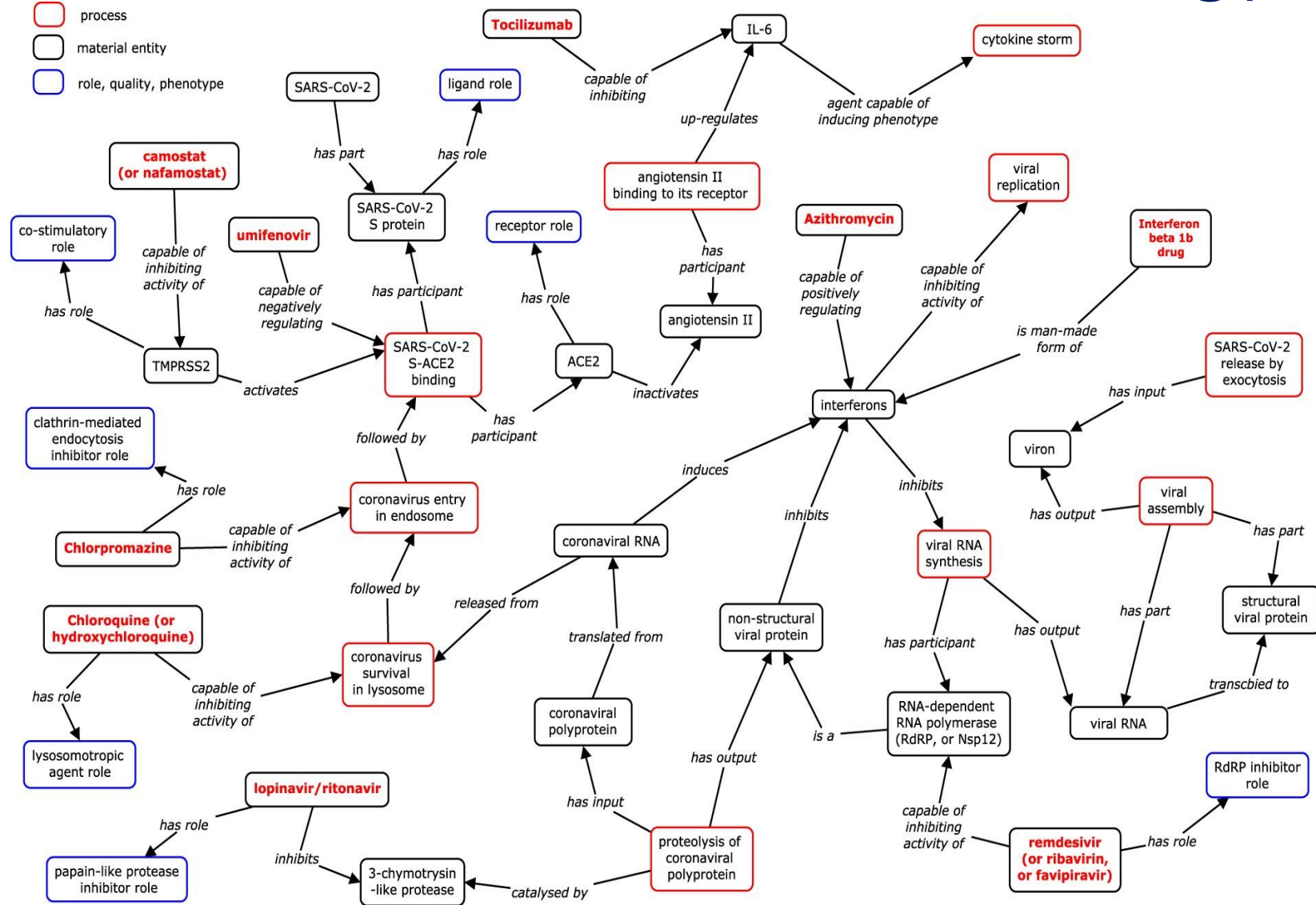


What is an ontology and why use them?

An ontology is a **structured framework** that defines concepts, entities, and their relationships within a **specific domain of knowledge**. It provides a **shared vocabulary** and a set of rules to organize information, enabling systems, people, and organizations to communicate and reason about that **knowledge in a consistent, meaningful way**.



CIDO: coronavirus infectious disease ontology



Our strategic priorities



Delivering Data-Driven
Value at Scale



Harnessing AI to
Expedite R&D



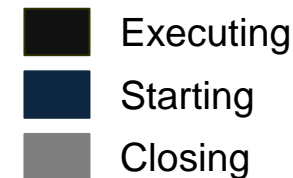
Accelerate Use of
RWD/RWE



Sustainability Driven
R&D



Activities Aligned with Strategic Priorities



Delivering Data-Driven Value at Scale

Ontologies

- IDMP-O
- Clinical Trials Operations
- CMC Process Chemistry
- Pharma General Ontology

Data Enrichment

- DataFairy Bioassay Annotation
- FAIR Implementation
- FAIR Community

Integration and Management of Data, Ontologies and Standards

- Methods HUB
- In-vitro Pharmacology
- Data Governance Community
- GSRS Consortium

Harnessing AI to Accelerate R&D

Algorithms

- Large Language Models in Life Sciences

Best Practices

- NLP Use Case Database
- AI/ML Community

Sustainability Driven R&D

- Carbon Footprint for Decentralized clinical trials
- Diversity & Inclusion in STEM Training / Community
- Digital Transformation & Change Management Community

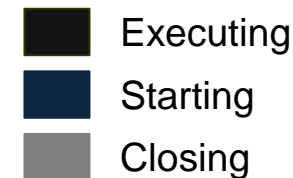
Accelerate Use of RWD/RWE

- Social Media for RWD

Other Projects And Communities

- Microbiome-Mediated Drug Metabolism Database
- Controlled Substance Compliance & Shipping Community
- User Experience Life Sciences Community
- Labs of the Future Community
- Quantum Computing Community

Activities Aligned with Strategic Priorities



Delivering Data-Driven Value at Scale

Ontologies

- IDMP-O
- Clinical Trials Operations
- CMC Process Chemistry
- Pharma General Ontology

Data Enrichment

- DataFairy Bioassay Annotation
- FAIR Implementation
- FAIR Community

Integration and Management of Data, Ontologies and Standards

- Methods HUB
- In-vitro Pharmacology
- Data Governance Community
- GSRS Consortium

Harnessing AI to Accelerate R&D

Algorithms

- Large Language Models in Life Sciences

Best Practices

- NLP Use Case Database
- AI/ML Community

Sustainability Driven R&D

- Carbon Footprint for Decentralized clinical trials
- Diversity & Inclusion in STEM Training / Community
- Digital Transformation & Change Management Community

Accelerate Use of RWD/RWE

- Social Media for RWD

Other Projects And Communities

- Microbiome-Mediated Drug Metabolism Database
- Controlled Substance Compliance & Shipping Community
- User Experience Life Sciences Community
- Labs of the Future Community
- Quantum Computing Community

Ideas Aligned with Strategic Priorities

Delivering Data-Driven Value at Scale

Ontologies

- Instrument Equipment Ontology
- Experiment & Assay Ontology (BAO)

Data Enrichment

- Complex Bioassay Data Exchange
- Regulatoria (IVP for Analytical)
- CDSIC for early/preclinical R&D

Integration and Management of Data, Ontologies and Standards

- Chemical Exchange File Format Community
- Standardized bioinformatics (NGS) computational pipelines
- Standard API for Semantic Data

Harnessing AI to Accelerate R&D

Algorithms

- In-silico testing
- Multi-modal LLMs for Webinars
- FDA2.0/Digital Twins for Animal Reduction
- AI for Drug Repurposing

Best Practices

- Model description vocabulary
- Utilization of AI/ML results in regulatory filings

Sustainability Driven R&D

- In Vitro NAMs Data Standards
- In Vivo metadata vocabulary

Accelerate Use of RWD/RWE

- RWD Ontology
- RWD and the impact of AI

Other Projects And Communities

- Regional Restrictions on Shipping & Data Exchange
- Blockchain technology for quality assurance

Ideas Aligned with Strategic Priorities

Delivering Data-Driven Value at Scale

Ontologies

- **Instrument Equipment Ontology**
- **Experiment & Assay Ontology (BAO)**

Data Enrichment

- Complex Bioassay Data Exchange
- Regulatoria (IVP for Analytical)
- CDSIC for early/preclinical R&D

Integration and Management of Data, Ontologies and Standards

- Chemical Exchange File Format Community
- Standardized bioinformatics (NGS) computational pipelines
- **Standard API for Semantic Data**

Harnessing AI to Accelerate R&D

Algorithms

- In-silico testing
- Multi-modal LLMs for Webinars
- FDA2.0/Digital Twins for Animal Reduction
- AI for Drug Repurposing

Best Practices

- Model description vocabulary
- Utilization of AI/ML results in regulatory filings

Sustainability Driven R&D

- In Vitro NAMs Data Standards
- In Vivo metadata vocabulary

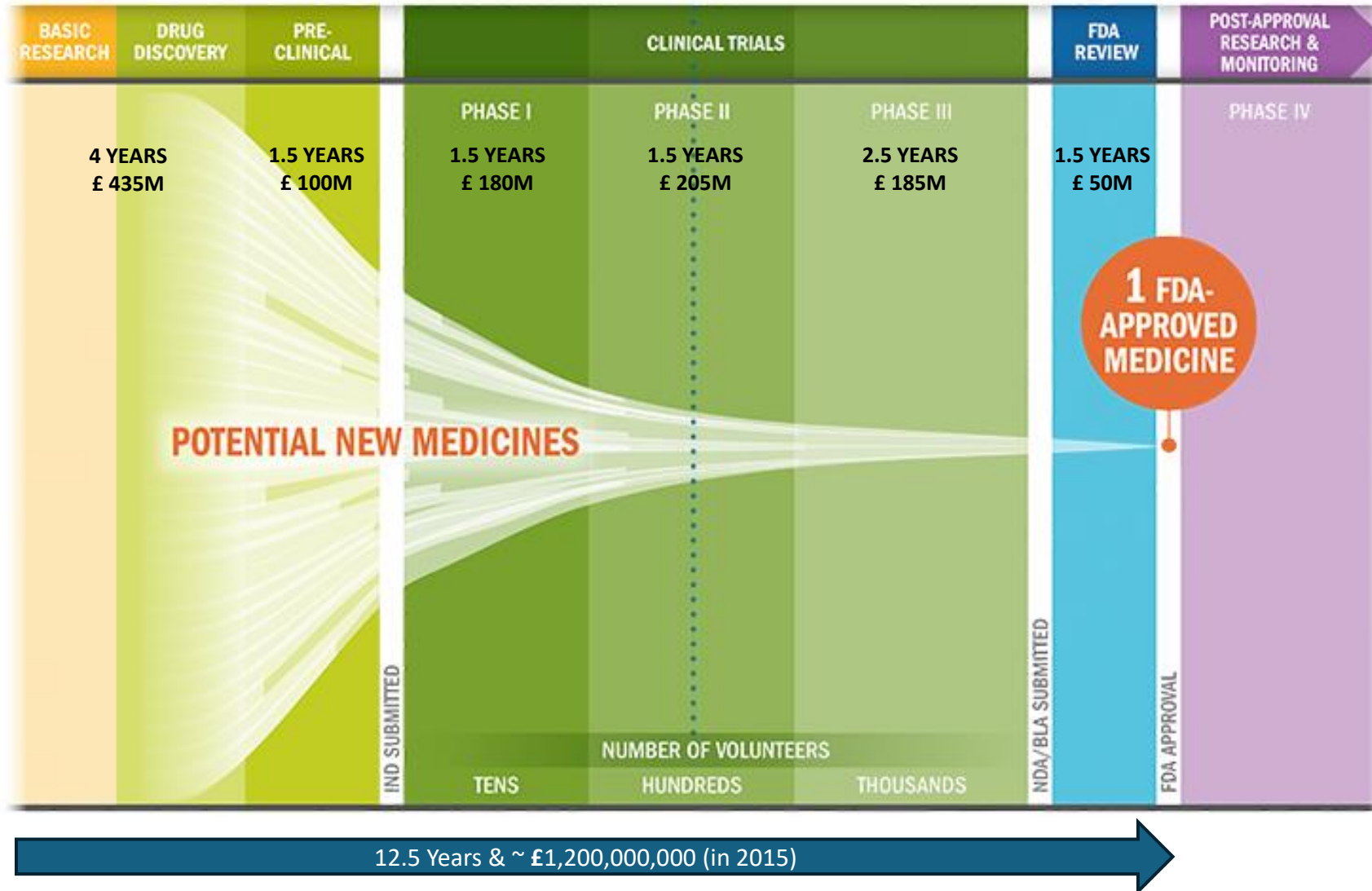
Accelerate Use of RWD/RWE

- **RWD Ontology**
- RWD and the impact of AI

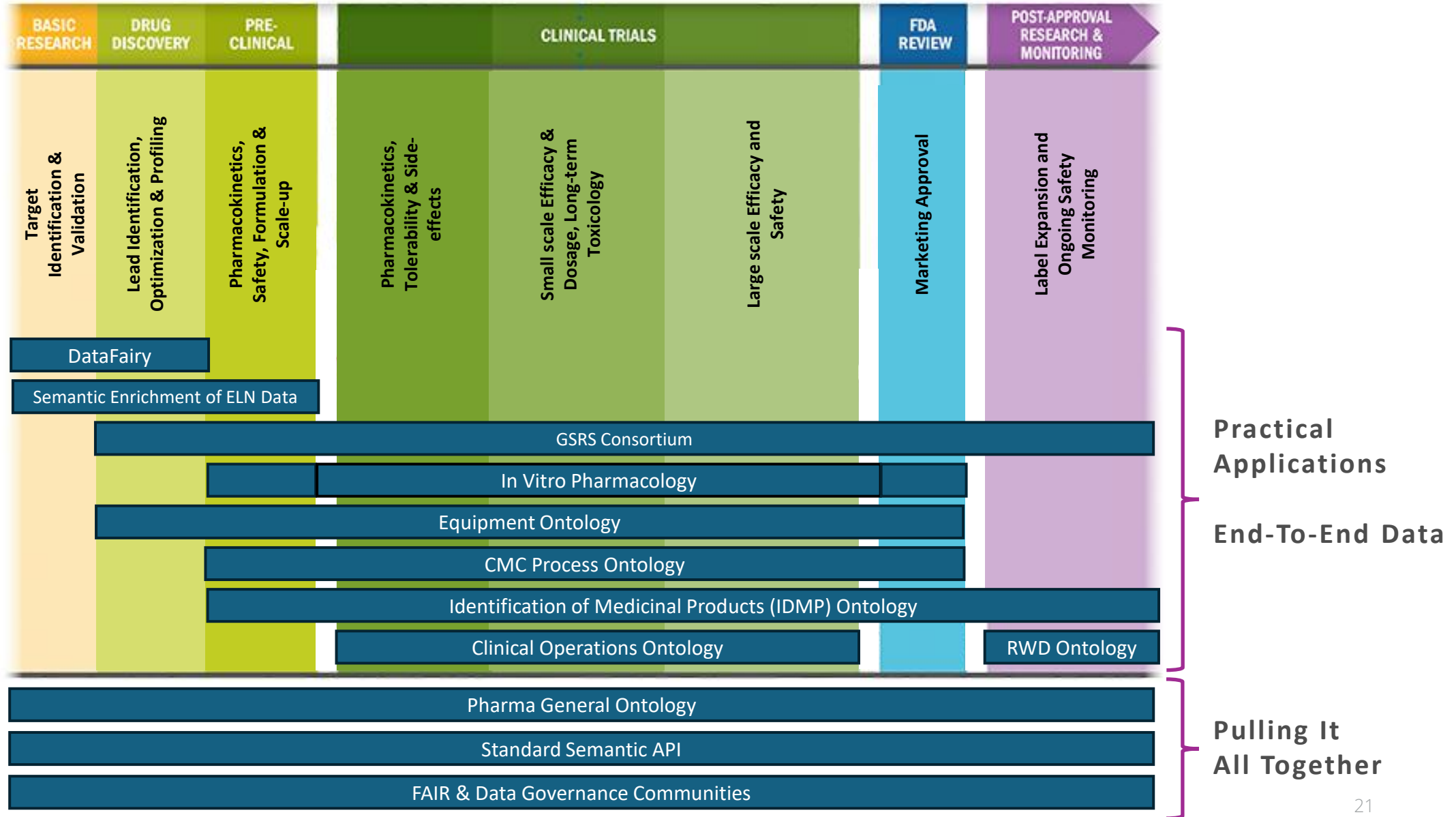
Other Projects And Communities

- Regional Restrictions on Shipping & Data Exchange
- Blockchain technology for quality assurance

Drug Discovery & Development Process

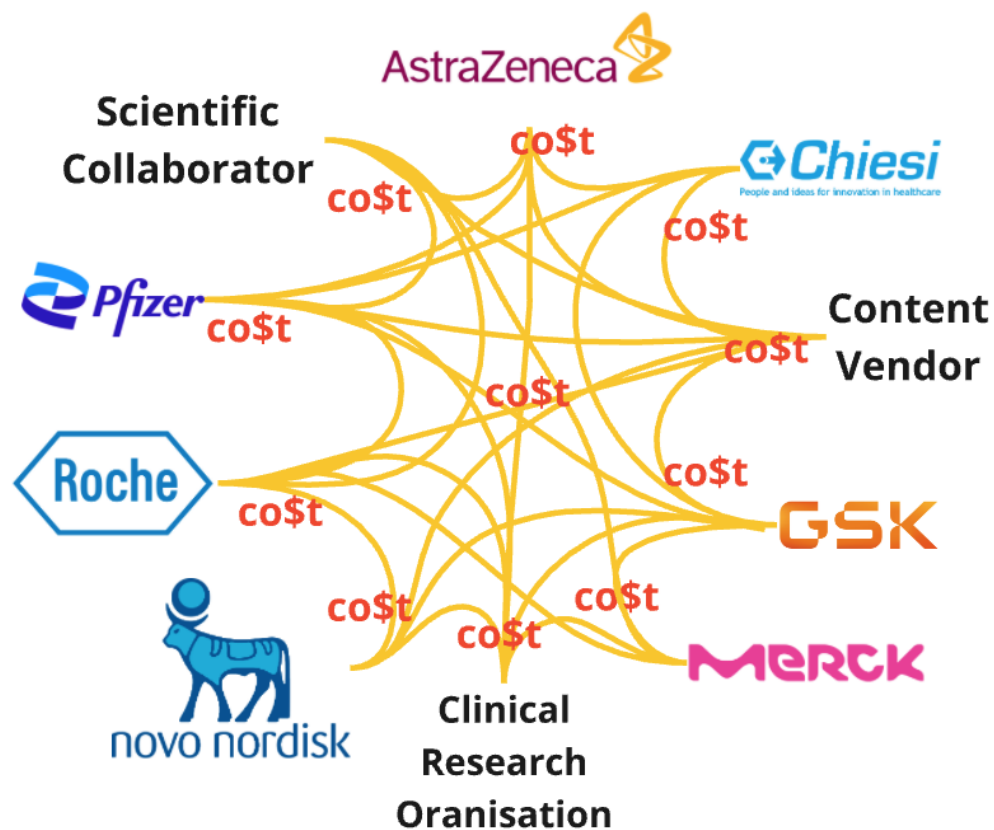


Drug Discovery & Development Process

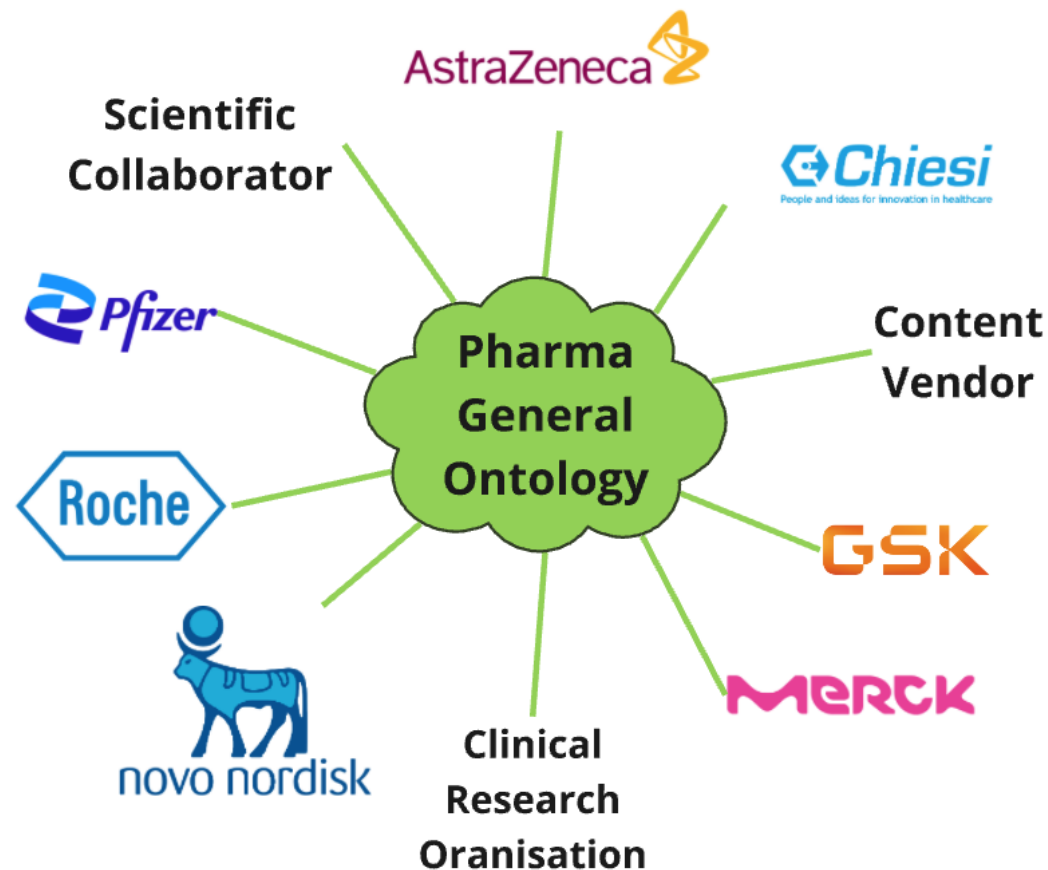


Pharma General Ontology (PGO)

Current situation

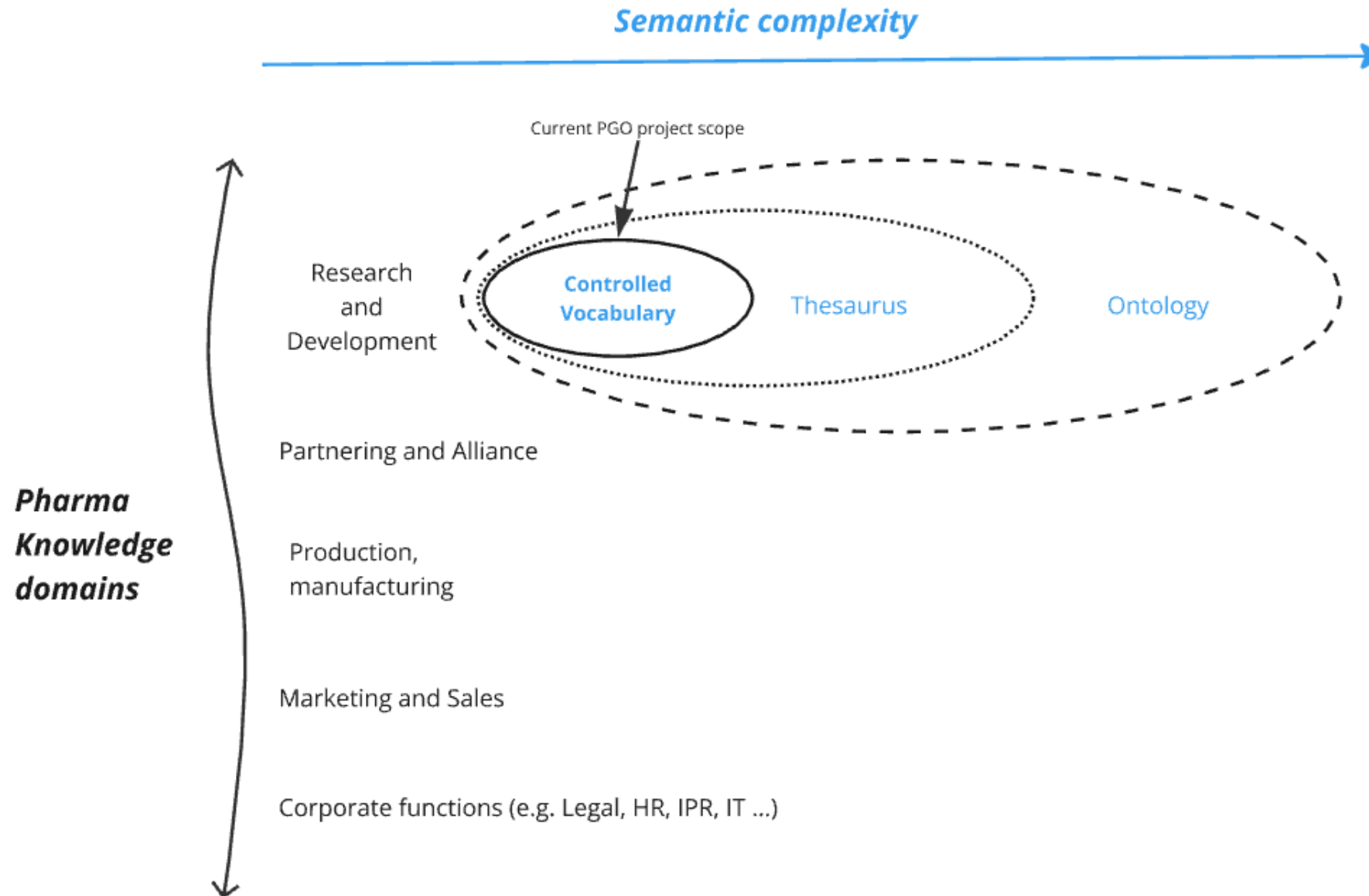


Future desired state



Pharma General Ontology (PGO)

Current & Future Scope



IDMP- Ontology (2024)

PM: Aditya Tyagi

Start: 1-Jan-2024

End: 31-Dec-2024

Next Phase: 1-Jan-2025

The Challenge

IDMP is a set of ISO Standards for the **ID**entification of **M**edicinal **P**roducts that is being included in EMA regulations to ensure consistent Identification, Documentation and Exchange of medicinal product information.

However, implementations of the standard across companies and regions are not standard, diluting expected benefit. Pistoia Alliance is bringing the industry together to build a common Ontology (IDMP-O) to implement the IDMP standards.

The Value Proposition: Interoperability across Global Regulatory & Healthcare Communities

Objectives:

1. **Support for implementation of IDMP-O in pharmaceutical companies**
2. **Extend IDMP-O coverage of IDMP** according to “Guided-Depth Principle”, i.e., based on pharma implementation feedback and new use cases (EMA Shortage PoC, Batch tracking and Automation)
3. **Facilitate industry adoption of IDMP-O** through professional communication, trainings, events and an Authorized Partner Program with consultancy firms and software vendors
4. **Increase integration aspects of IDMP-O:** e.g., primarily ISA-S88, secondary with Clinical, FHIR based on pharma implementation needs and new use cases
5. **Transition to Maintenance:** Build a long-term funding & governance operating model

Project Funders



Clinical Operations Ontology (2024)

PM: Aditya Tyagi

Start: 1-Nov-2023

End: 31-Aug-2024

Next Phase: 1-Sept-2024

The Challenge

Most clinical trial protocols are written with key information in free text. While standards exist for the codification of protocol information, no mechanism exists to transform protocol-level information into clinical operations executables.

Will an ontology streamline the processes linking protocol-level instructions with clinical operations executables?

The Value Proposition: Improve Interoperability & Efficiency of Clinical Trials Task Execution

Goal:

- Develop vendor agnostic, open-source clinical operations ontology supporting the collection, analysis and exchange of ClinOps data.
- Initial focus is to add a semantic layer to USDM 2.0 and additional operational data from the clinical protocol

Deliverables include:

- Use cases and competency questions to demonstrate value
- Minimal viable, vendor-agnostic open-source process clinical operations ontology

Project Funders (under discussion for phase 2)



Pharmaceutical CMC Process Ontology (2024)

PM: Birthe Nielsen

Start: 1-Oct-2023

End: 31-Aug-2024

Next Phase: 1-Sept-2024

The Challenge

Chemical development, has been supported by the ISA 88 standard for decades. A lot has changed in that time, and a more flexible, comprehensive representation of process data is needed.

Phase 1 (PoC) has shown that a contemporary ontology will provide the data representation that is needed to facilitate data integration, data exchange and data insights.

The Value Proposition: Support Today's Data Science, Integration & Exchange Needs

Goal: Develop vendor agnostic, open-source process chemistry ontology supporting today's data science and exchange needs

Benefits:

- Easier exchange of data between collaborators
- Machine readable data, and
- Improved data accessibility for machine learning & AI applications.

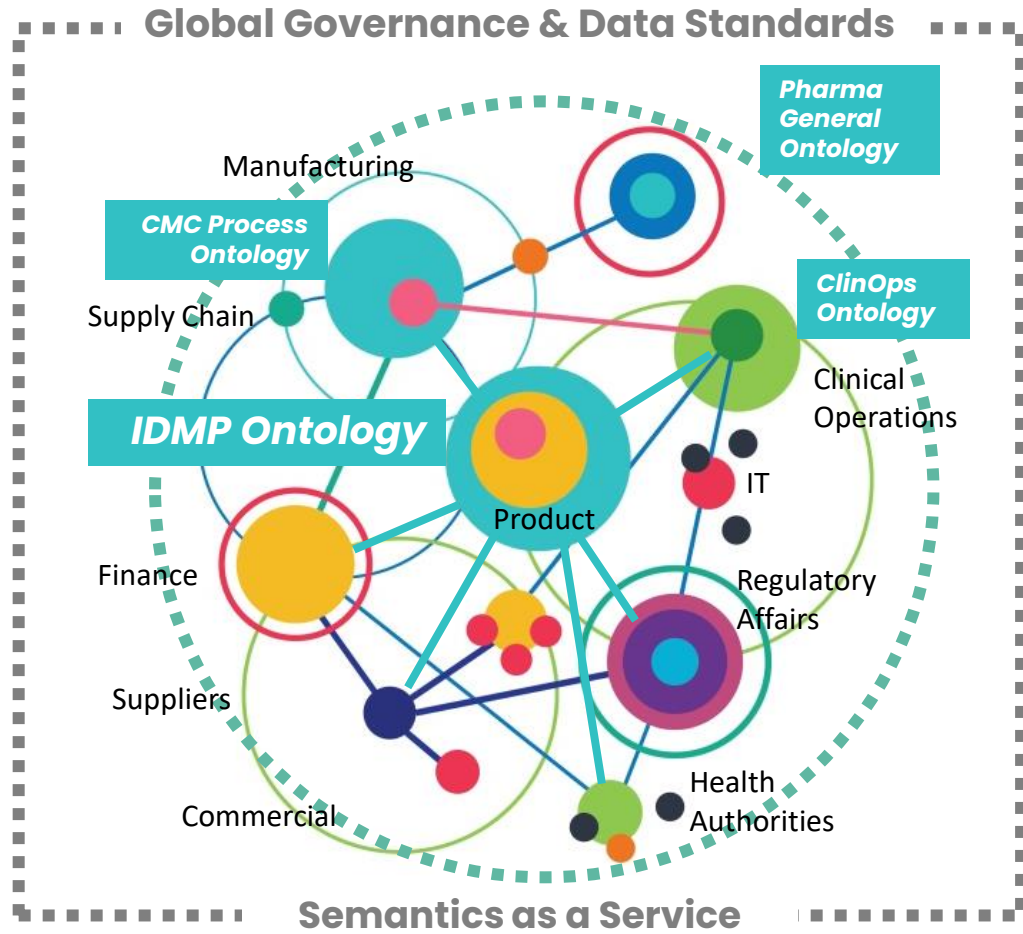
Deliverables:

- Minimal viable, vendor-agnostic process chemistry ontology
- Process Definition, Process Monitoring, and Comparative Analysis use cases for APIs and drug products for both biologics and synthetics

Project Funders



Pistoia Alliance Semantic Network



A sustainable portfolio of actionable lifesciences ontologies & semantic resources coordinated by an in-house Ontologist to ensure interoperability.

IDMP Ontology acts as the Backbone and the Pharma General Ontology provides an overall Rosetta Stone for the industry.

Upcoming Ontology Projects

Instrumentation & Equipment Ontology

PM: Birthe Nielsen

Start: Jan-2025

End: Dec-2025

Next Phase: Jan-2026 (TBC)

The Challenge

A lack of an ontology and standard vocabularies in instrument and equipment inventories makes the categorizing and querying of these systems difficult

This has led to challenges in building efficient data pipelines, matching process/analytical method requirements with receiving site capabilities and ensuring cyber security and efficient maintenance of the instrument & equipment fleet.

The Value Proposition:

Streamline data management to improve efficiency and security, and facilitate technology transfer

Work with both pharmaceutical companies and instruments/equipment vendors to create an ontology covering instrument & equipment inventory to provide a structured framework for organizing knowledge.

Benefits include improved efficiency in data management, cost savings from avoiding redundant purchases, improved budget forecasting, better asset allocation and utilization, enhanced compliance & cyber security, streamlined site impact assessment and tech transfer, and facilitated management of OT metadata.

Ontology will need to fully integrate with Pharma CMC Process Ontology.

Project Supporters



Experiment & Assay Ontology

PM: TBD

Start: Jan-2025

End: Dec-2025

Next Phase: Jan-2026 (TBC)

The Challenge

The current ontologies available to support annotation of experiments and assays (BAO, AFO) have not been updated for some time and are not yet sufficient to fully enable experiment data interoperability and their reusability.

In particular there is a need for a better model of the relationship between protocols, experiments and results and the concepts of measurements which lies at the core of experiments.

The Value Proposition:

Update and Expand the BAO Ontology to meet current Pharma needs

- Work with the current BAO team, and existing Pistoia Alliance projects, to upgrade the ontology to enable accurate (i.e., correct) interpretation of experimental results by both humans and machines.
- The initial proposal is to establish a formal definition and a generic model for measurements, experimental roles & to eventually to clarify the concept of “experiments” of all types.
- The project will produce an updated, extensible, ontology capable of supporting experiments and their results, allowing data interoperability, aggregation (as appropriate) and reusability.
- Ongoing support and sustainability of the ontology will also be a key goal of the project.

Project Supporters

 Galápagos



Standard Semantic API

PM: TBD

Start: Oct-2024

End: Dec-2025

Next Phase: Jan-2026 (TBC)

The Challenge

FAIR data strategies, drive the creation and deployment of data terminologies and ontologies to achieve data interoperability.

For data to be “born FAIR”, systems e.g. ELNs and LIMS, need to surface terminologies at the time of authoring.

While these systems support controlled terminologies, they are often closed and integrate poorly with ontology and terminology systems (OWS).

The Value Proposition:

Provide a base specification for API access to semantic data

The goals are to:

- Develop a common API protocol for a common set of operations relating to ontology and terminology lookup
- Provide a conventional API based on the RESTful style architecture and abstract away from any semantic web standard such as OWL or SKOS to enable a diversity of terminology types and use cases
- Eliminate or minimize need to develop and maintain complicated middleware

Project Supporters



The Challenge

Key to an effective use of RWD is the ability to find such data, understand their characteristics and their fitness for purpose. While standards for the representation of RWD are being developed (e.g. OMOP), there is no standard for the metadata describing such datasets. Such metadata need to cover aspects related to collection process, biases, consent, standards and more. With thousands of sources available, the lack of consistent metadata hampers findability and even usability of RWD.

The Value Proposition: Develop a metadata standard to describe RWD sources

The goal of the project is to develop a metadata standard to describe RWD sources, that can be machine-actionable and enable semantic-based findability and integration.

Such metadata scheme will provide a best and shared way to describe the content, quality, focus and characteristics of a RWD dataset in a way that would enable the understanding of which source is best to answer specific scientific questions.

Project Supporters :

Budget under evaluation following scoping of the project



General Ontologies Training

11-Oct-2024 to 8-Nov-2024

Goal: Provide a general understanding of ontologies and their business value



Webinars

Introduction and Business Value of Ontologies

From Taxonomies to Ontologies and Knowledge Graphs

Usage of Ontologies: Use Cases from the Pharmaceutical Industry

Ontologies for Pharma and Life Sciences Leaders

Increasing Adoption to Realize Value

Target Audience:

- Pharma business leaders & budget holders,
- Technology leaders and scientists

Format:

- Five weekly live 1-hour webinars online
- Followed by on-demand access

Speakers (TBC)

Christine Memmott, Vertex

Martin Romacker, Roche

Andrea Splendiani, IQVIA

Sheila Elz, Bayer

Joshua Valdez, Novo Nordisk

Birgit Meldal, Pfizer

Jane Lomax, Scibite

Rajaram Kaliyaperumal, J&J

Berenice Wulbrecht, Ontoforce

Recommended reading

**Ontologies 101- using wine as an example!
(Stanford University)**



**Ontologies4Chem: the landscape of ontologies in
chemistry**





Thank you





**Ontologies in Pharma:
The landscape, pre-competitive
development, and use-cases**

Becky Upton
Pistoia Alliance
President