Importance of Good Data Management

Experience from a mid-sized biotech

Nicolas Triballeau, Ph.D.

Director, Drug Discovery Chemistry

revvity signals

November 18th, 2025

Context

- European Biotech created in 1999
  - Target discovery platform
  - Drug discovery, mostly small molecules
  - By 2018 : 725 employees and millions of experimental results generated

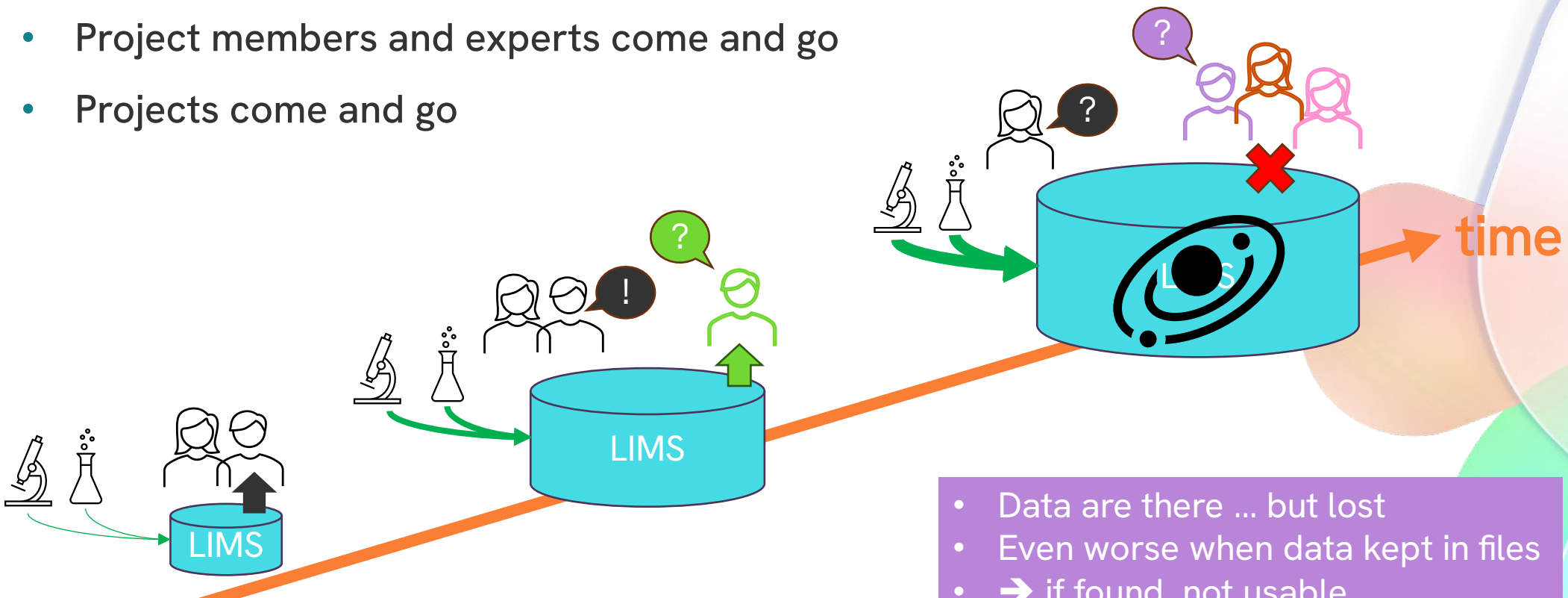
- Data platform
  - Designed as a LIMS system, originally to support target discovery
  - On premise Oracle database with PHP coded interface for querying

revvity signals

November 18th, 2025

Situation : a data “black hole”

- Project members and experts come and go
- Projects come and go



- Data are there … but lost
- Even worse when data kept in files
- ➔ if found, not usable
- ➔ re-do experiments (ttt, $$$)

revvity signals

November 18th, 2025

FAIR data principles

Findable

- (Meta)data have **unique & persistent identifier** (e.g., DOI)
- **Rich metadata** describing the data
- Metadata includes data identifier
- Registered in searchable resource/repository

Accessible

- Retrievable via standardized protocol (HTTP, FTP)
- Protocol is open, free & universally implementable
- Authentication & authorization when needed
- Metadata remains accessible even when data is not

Interoperable

- Formal, shared, broadly applicable **language**
- **Vocabularies** following FAIR principles
- Qualified references to other data/metadata
- Machine-readable & exchangeable formats

Reusable

- **Rich description** with accurate attributes
- Clear & accessible usage license
- Detailed provenance information
- Meets domain-relevant community standards

➔ Requires **semantics** = defining a common language in a community

revvity signals

The FAIR Guiding Principles for scientific data management and stewardship, Wilkinson et al., **2016** *Scientific Data*

Different levels of semantics elaborated with experts

strong semantics

weak semantics

Ontology
- description logic
- axioms
- rules

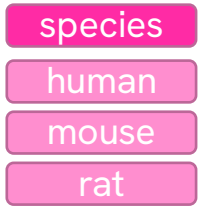
Conceptual Model

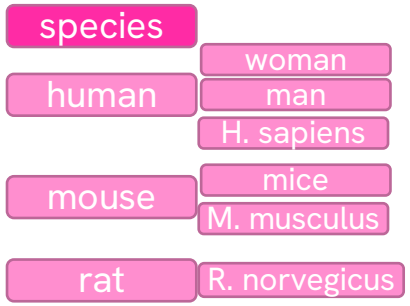
Taxonomy
- Formal *parent-child* relations ✔
- Generalization hierarchy ✔

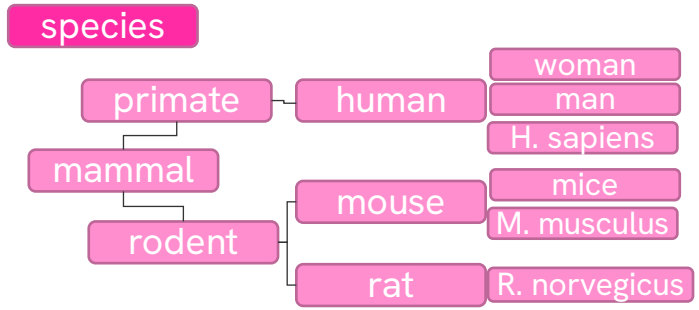
Thesaurus
- (Synonyms)

Informal Hierarchy

Controlled Vocabularies ("CV")





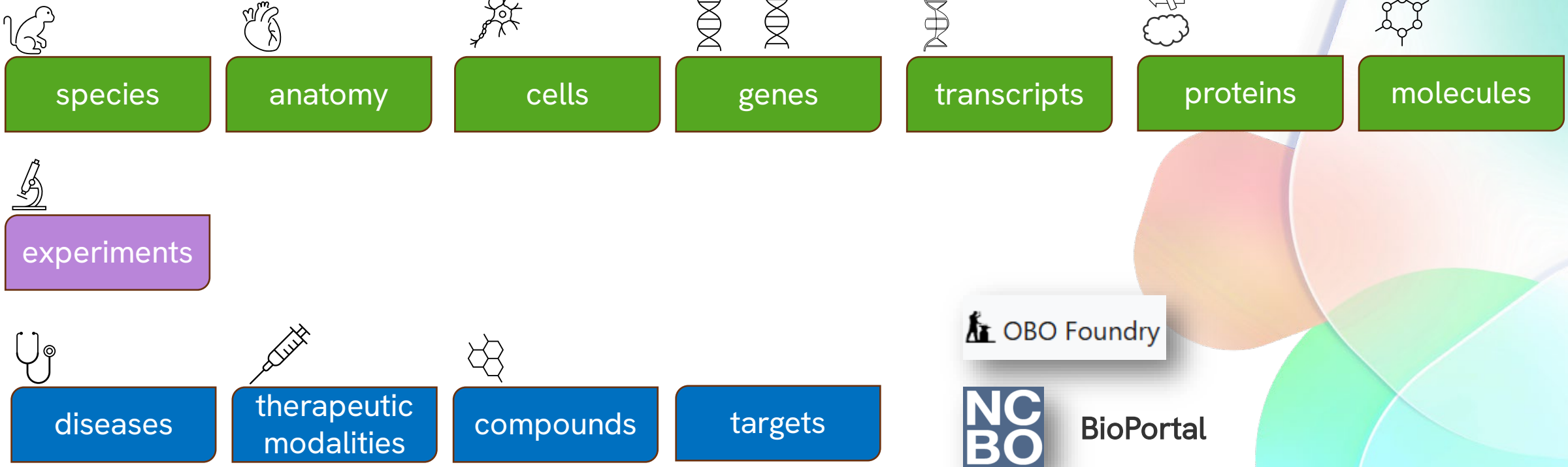


structural and syntactic interoperability

semantic interoperability

revvity signals

Ontologies in Life Sciences – Drug Discovery



OBO Foundry

NCBO BioPortal

OLS ONTOLOGY SEARCH

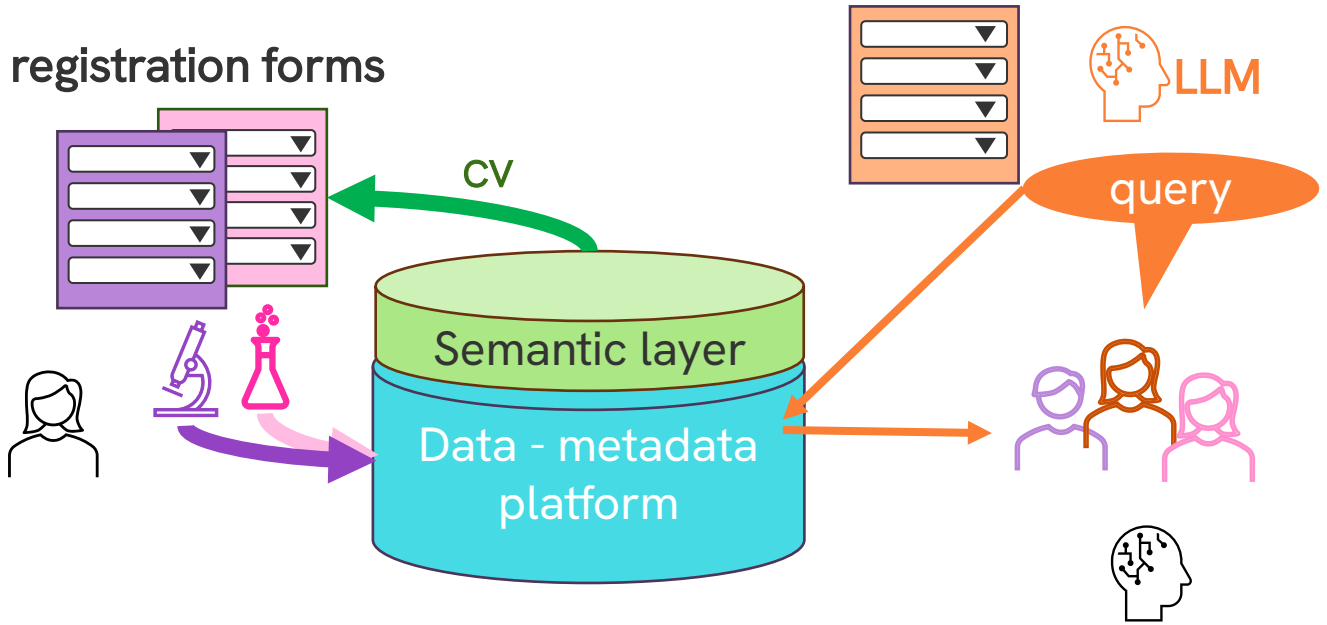
revvity signals

November 18th, 2025

Concretely

Data production

Data consumption



Data Analyses
Meta Analyses

AI/ML models/predictions

revvity signals

November 18th, 2025

Take home messages

- Follow the FAIR data principle
- When possible, annotate data by the experts who generate them
- Enforce use of terminologies, ideally supported by ontologies
- Benefits:
  - Capitalize on your data = your assets
  - Find them, reuse them, share them
  - Machines can properly interpret them ➔ AI/ML

revvity signals

November 18th, 2025