



Semantic Chemical Data Management

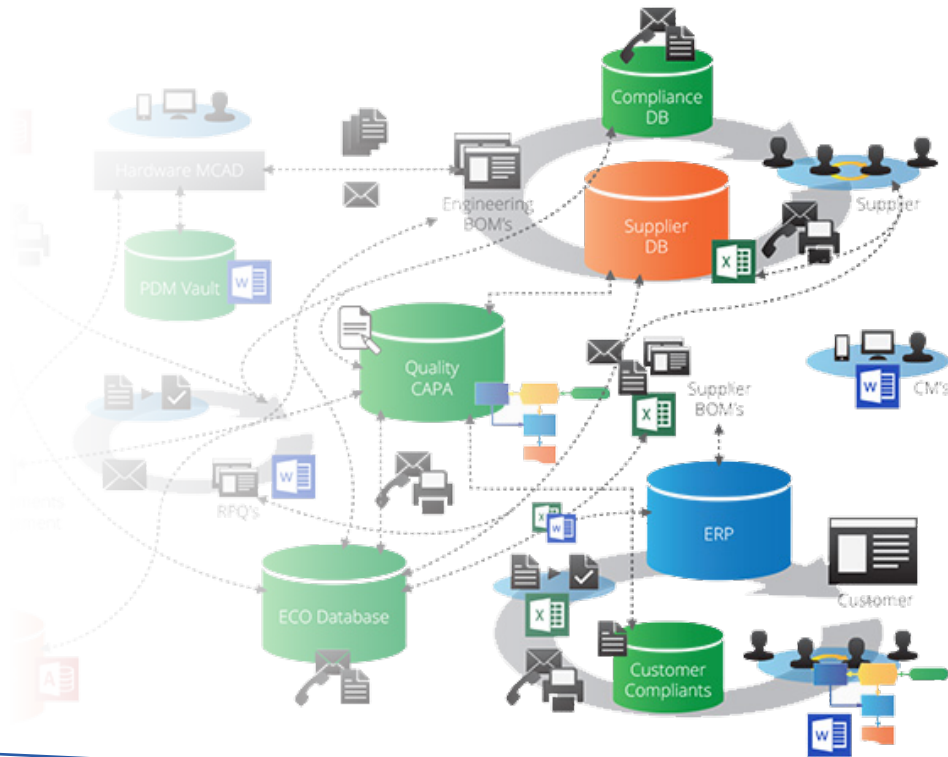
From FAIR to IDMP Compliance

Dr. Matthias Nolte

matthias.nolte@chemantics.com

Chemical Data Management Across the Pharmaceutical Value Chain

Regulatory Agencies (EMA, FDA, ...)



unFAIR data

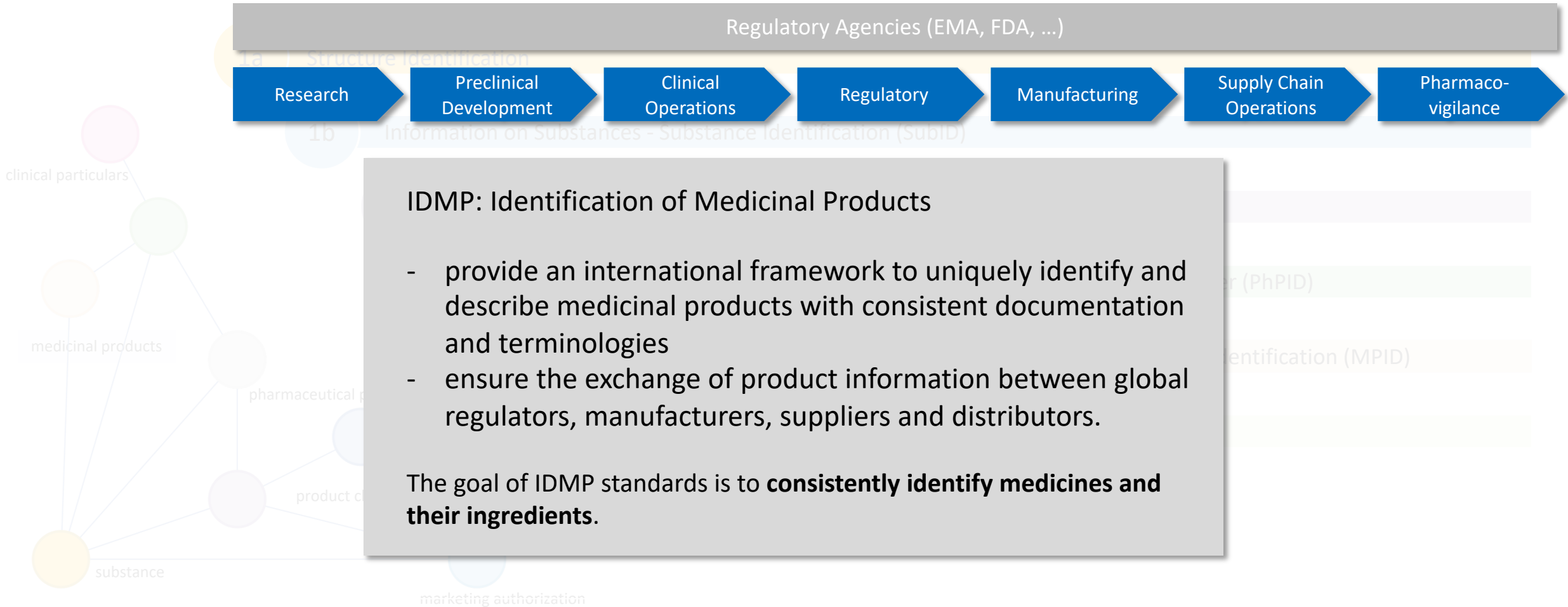
high integration efforts
risk of failure concerning submissions as
well as data and post market analysis

inconsistent management of chemical
structures and substance information

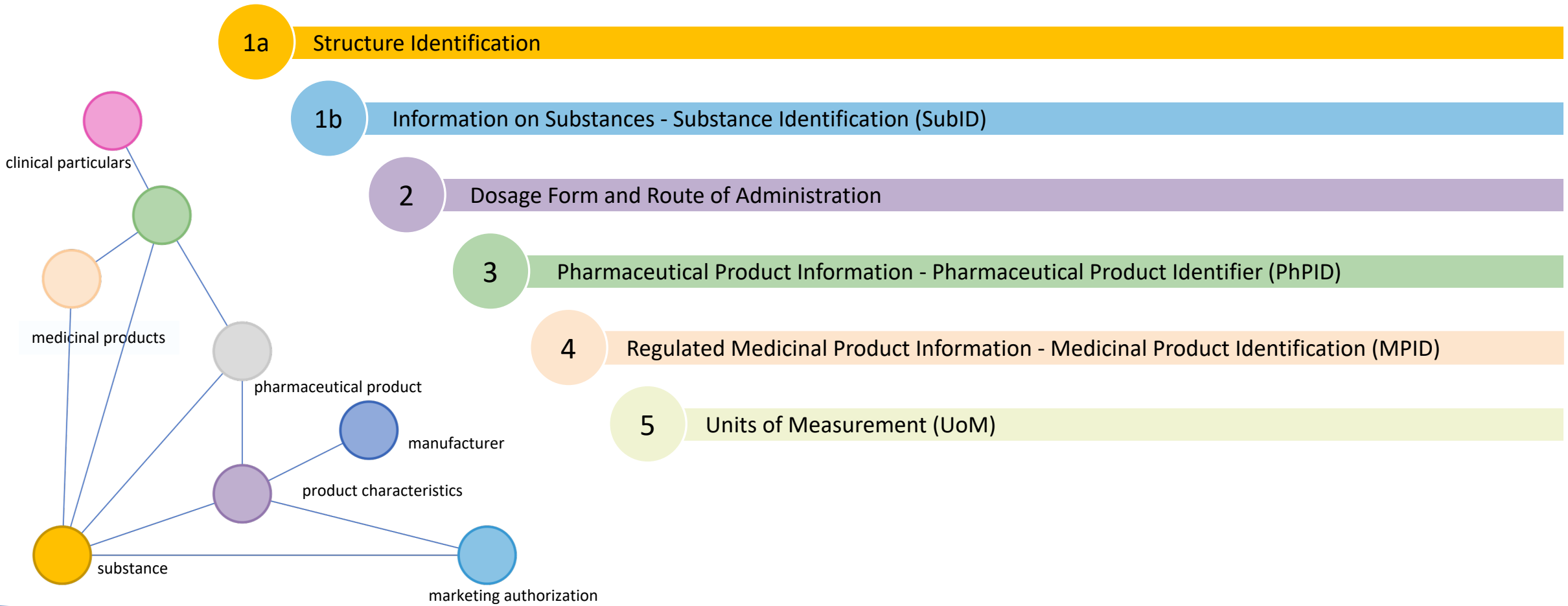
significant effort needed to identify
and link data across value chain

missing:
integration layer between data silos

The Pistoia IDMP Core Ontology Project



The Pistoia IDMP Core Ontology Project



The Pistoia IDMP Core Ontology Project

1a Structure Identification

1b Information on Substances - Substance Identification (SubID)

2 Dosage Form and Route of Administration

1a

FAIR chemistry data for chemical structure information and semantic alignment between structures

As a - *data analyst, data steward, data owner, ...*

I need to enrich my knowledge through association of internal and external data sources.

1b

Semantic alignment between structures and FAIRified substance information

As a – *regulatory affairs manager, data owner, business process owner, data steward, ...*

I need to manage substance related data consistently throughout the product life-cycle.

FAIR Chemistry (and FAIR Biology)



Registration

- generate normalized structure representation, PID
- generate structure/structure relationships

Organisation

- maintain source level database organization,
- categorize data sources via semantic relationships

Annotation

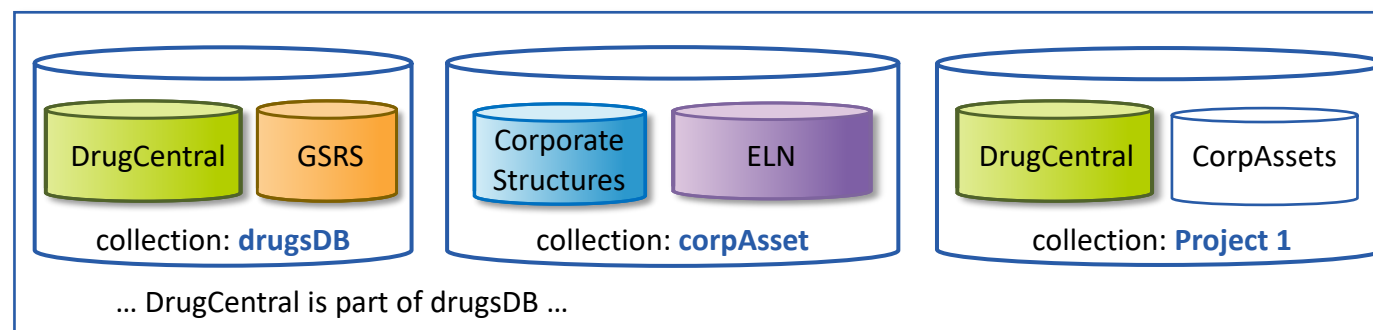
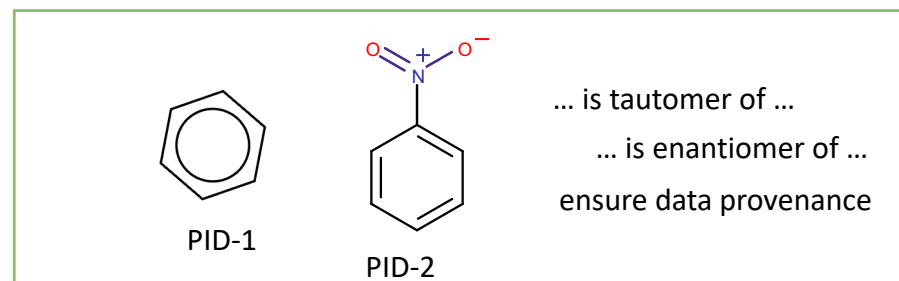
- annotate chemical entities through generation of multi-level abstractions and subsystem identification

Classification

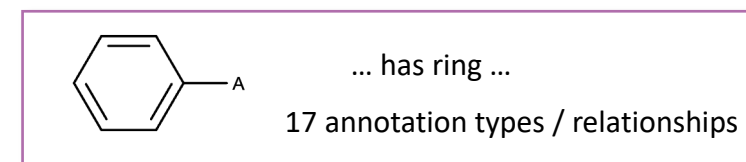
- classify chemical structures based on chemical structure classes

Integration

- API to access any data entity in the Chemantics eco system
- API to calculate structure based properties



... is an alcohol ...
... contains amino acid ...
... is a hydrocarbon ...
~ 2500 classifications

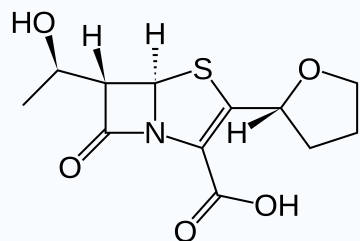


... create a new database collection
... register new compounds
... get calculated attributes
... search specific structure collections

Structure Graph: faropenem

Integration
access

Faropenem is an orally active beta-lactam antibiotic belonging to the penem group.[WIKI]



The **PubChem** database contains chemical molecules and their activities against biological assays.

The **DrugBank** database contains information on drugs and drug targets

The **Human Metabolome** Database (HMDB) contains detailed information about small molecule metabolites found in the human body.

The **Norman** Suspect List Exchange collects data relevant for environmental monitoring purposes.

PubChem

nber

DrugBank

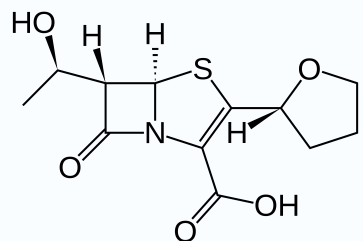
Human...

NORMAN

Structure Graph: faropenem



Integration
access

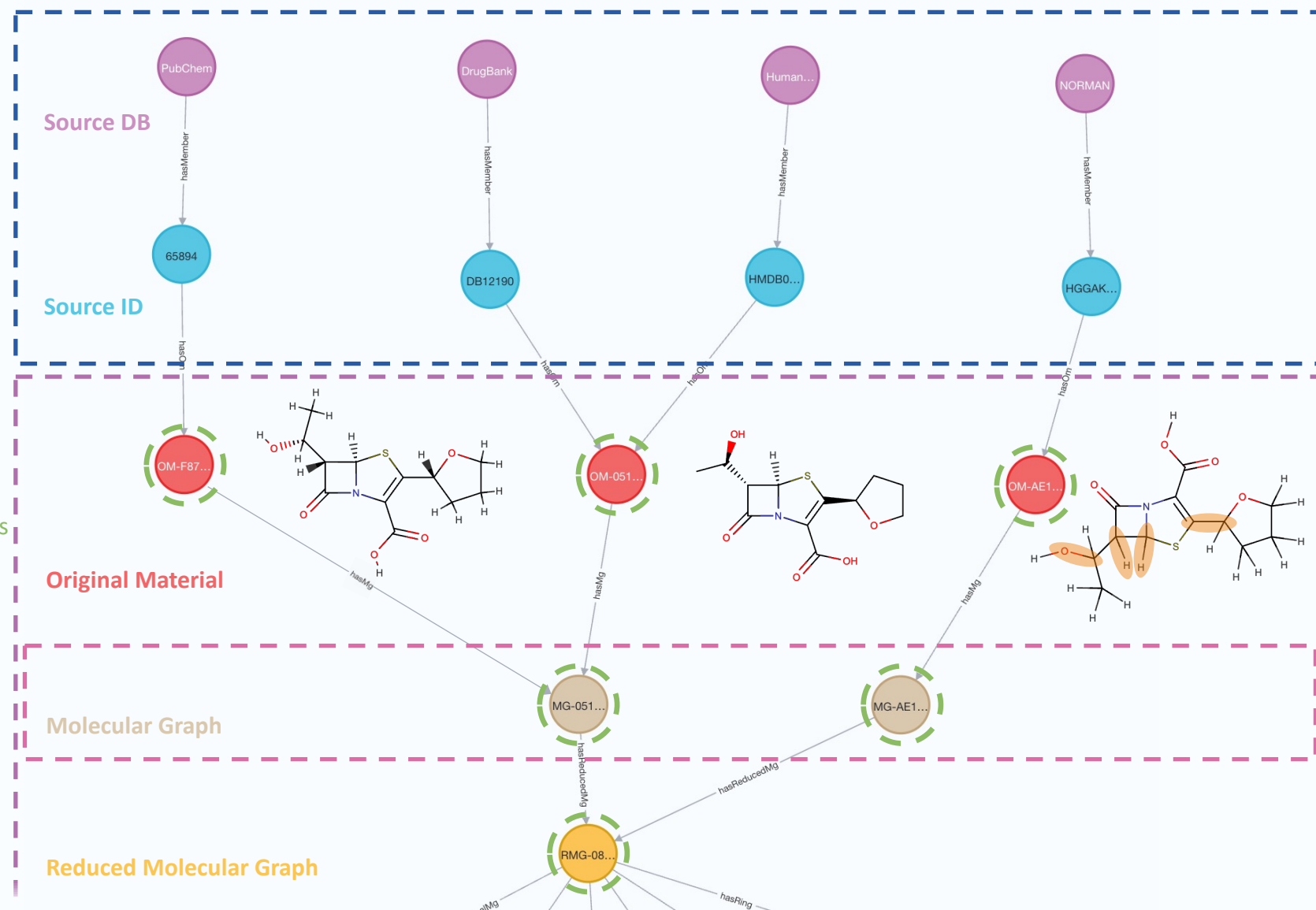


Organisation
provenance

Annotation
enrichment

Registration
structure PIDs

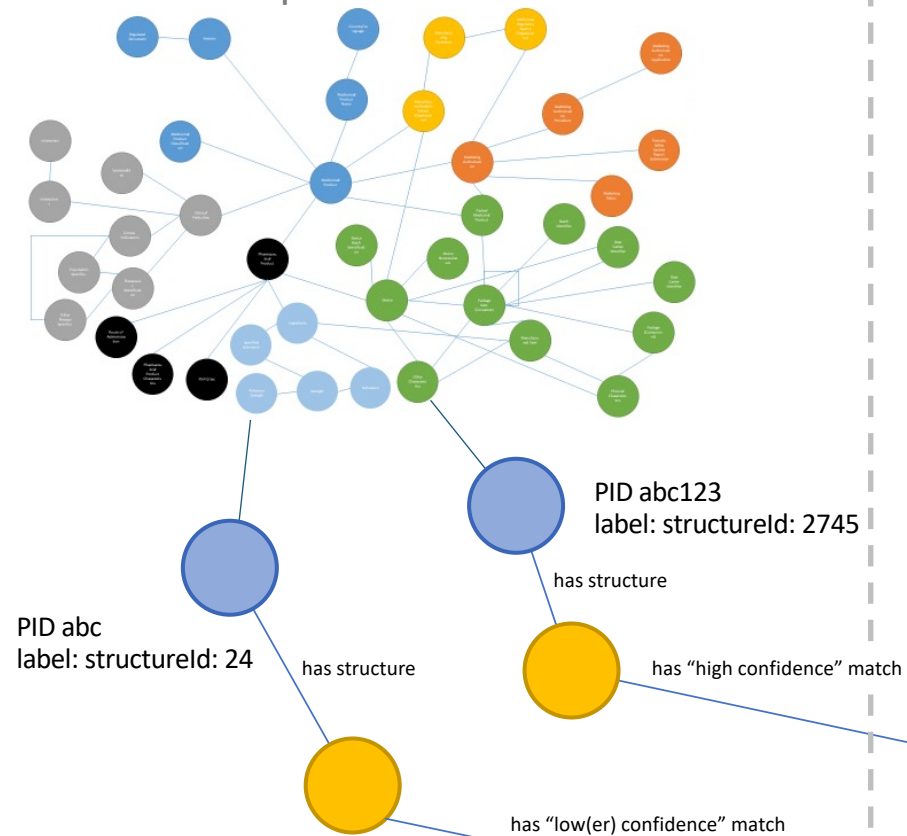
Classification
metadata



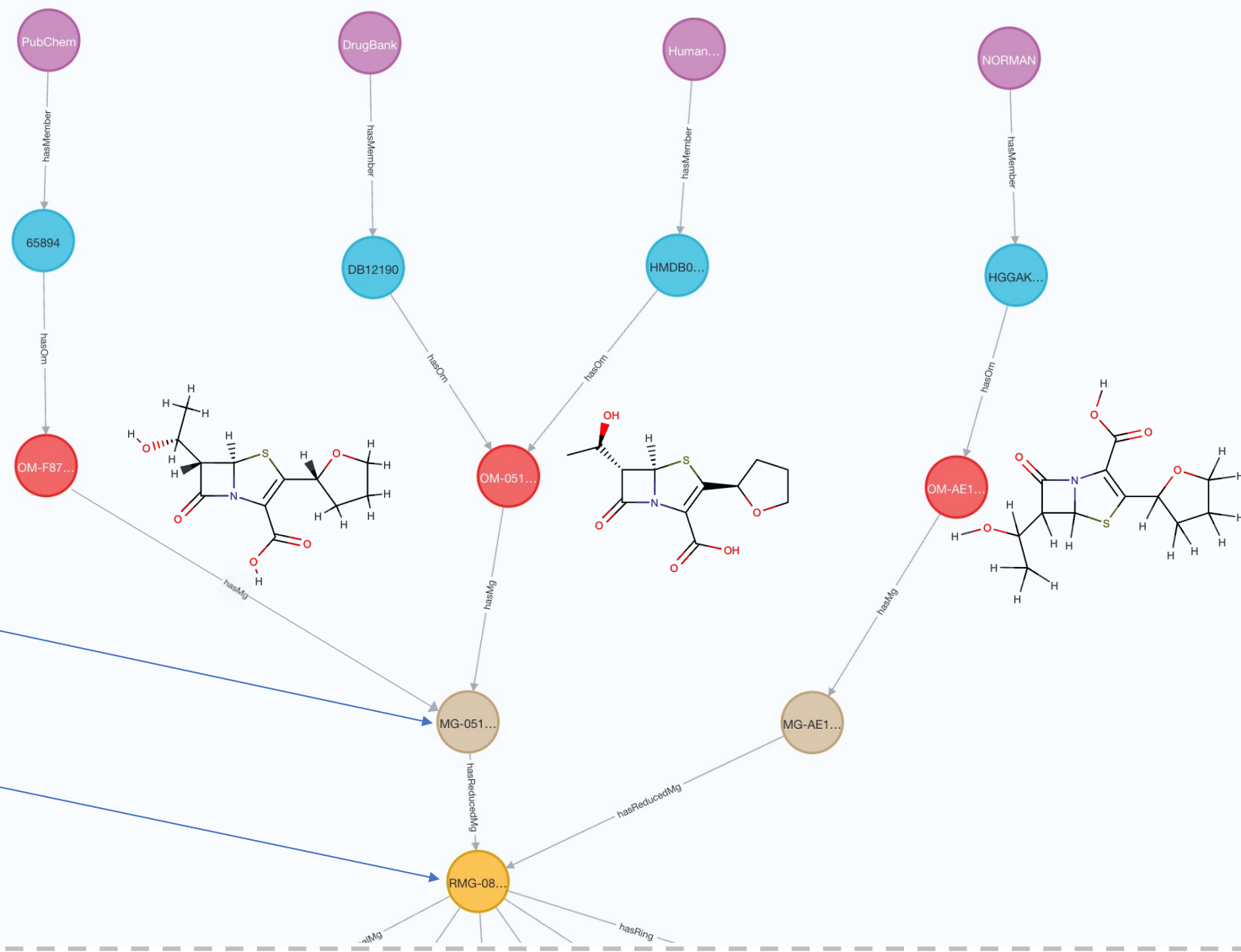
IDMP Core Ontology

How the Structure Graph Solves the Substance Identification Issue

Substance Graph



Structure Graph



Summary

- FAIR principles can be applied to chemistry representations in an automated and consistent manner through accurate identification and alignment of identical structures.
- Once FAIR Chemistry is implemented, the knowledge graph enables structure based data matching/linking with high accuracy.

Now, as a - *data analyst, data steward, data owner, ...*

I **can** enrich my knowledge through association of internal and external data sources.



Now, as a – *regulatory affairs manager, data owner, business process owner, data steward, ...*

I **can** manage substance related data consistently throughout the product life-cycle.



- The Pistoia's Alliance IDMP project integrates the combined capabilities
 - open core ontology
 - FAIR Chemistry

to support the implementation of IDMP standards as mandated by European Submissions office by 2023.



Questions