V.2.4

Allotrope Framework:
Semantischer Datenstandard für das Lab 4.0

Heiner Oberkampf, PhD
November 15th 2018, Duisburg
iuta 3. AnalytikTag

OSTHUS

WHO IS OSTHUS?

- Global organization
- 120+ employees and growing rapidly
- 16+ big pharma customers + many chemicals and lab-based companies.
- Allotrope Framework Architect
- Our approach technology:

Connecting data, people and organizations

OSTHUS

DATA AS AN ASSET

DIGITALIZATION

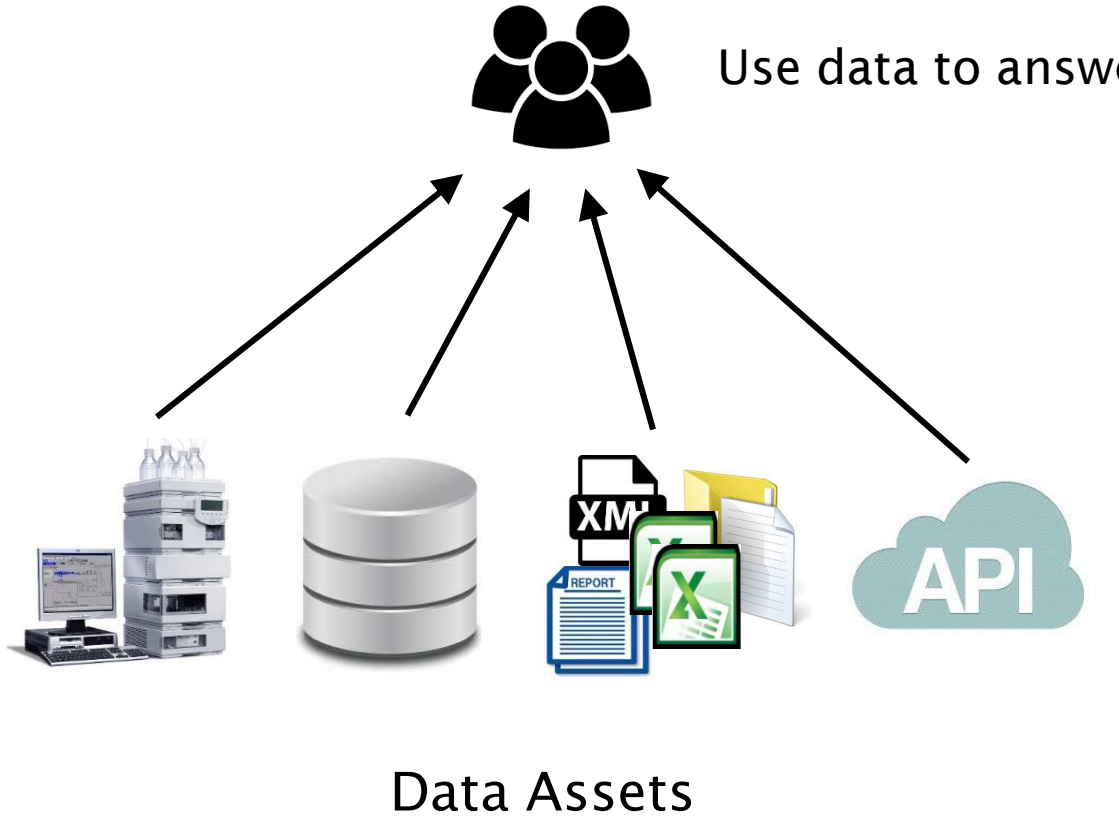
OSTHUS

Current Situation in the Lab

- increase in automation
- connected devices
- heterogeneous IT landscape
- increased complexity
- more data is generated
- stronger regulations
- high cost pressure

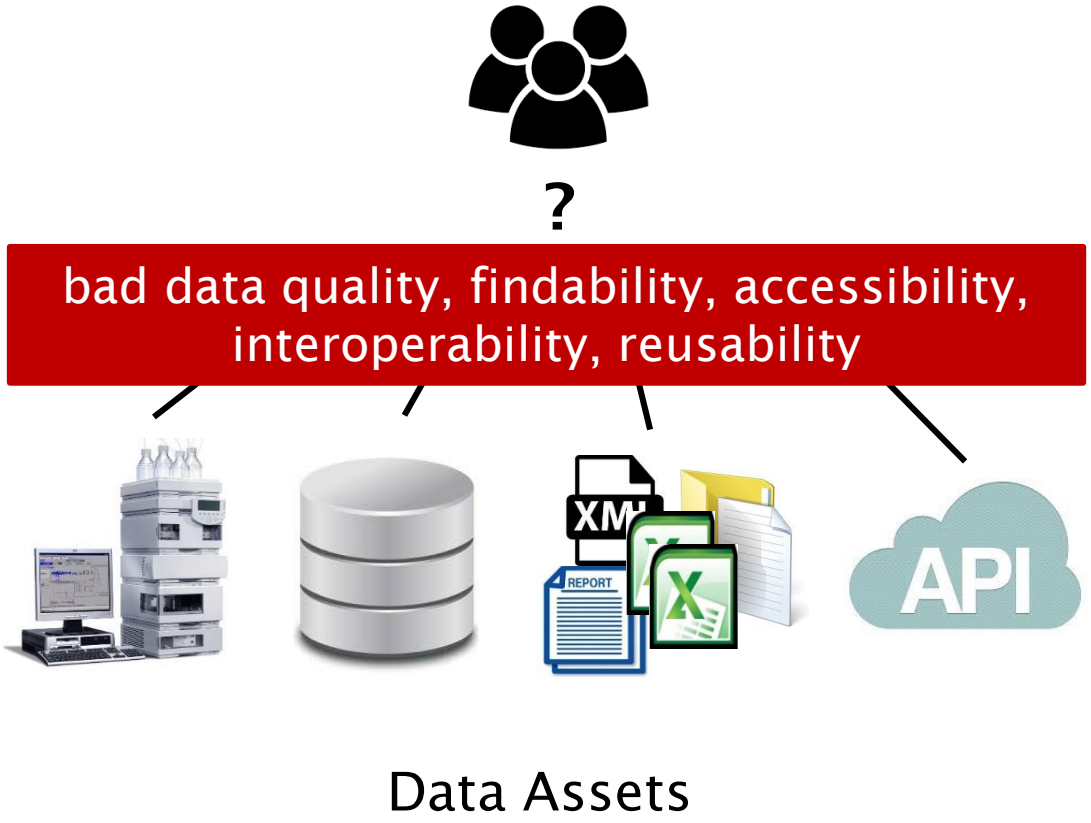
Digitalization is often still just “paper on glass”!

Value of Data is Realized by Usage in Decision Making and Insights



OSTHUS

Value of Data is Mostly Not Realized



Data Assets

"Only 3% of Companies' Data Meets Basic Quality Standards"

Harvard Business Review: https://hbr.org/2017/09/only-3-of-companies-data-meets-basic-quality-standards

"For most large enterprises, the root of this problem lies in years of treating the data generated by their operational systems as a form of exhaust rather than as a fuel to deliver great services, build better products, and create competitive advantage."

Database Trends and Applications: http://www.dbta.com/Editorial/Trends-and-Applications/The-Enterprise-Data-Debt-Crisis-123008.aspx

Finding Information is time-consuming

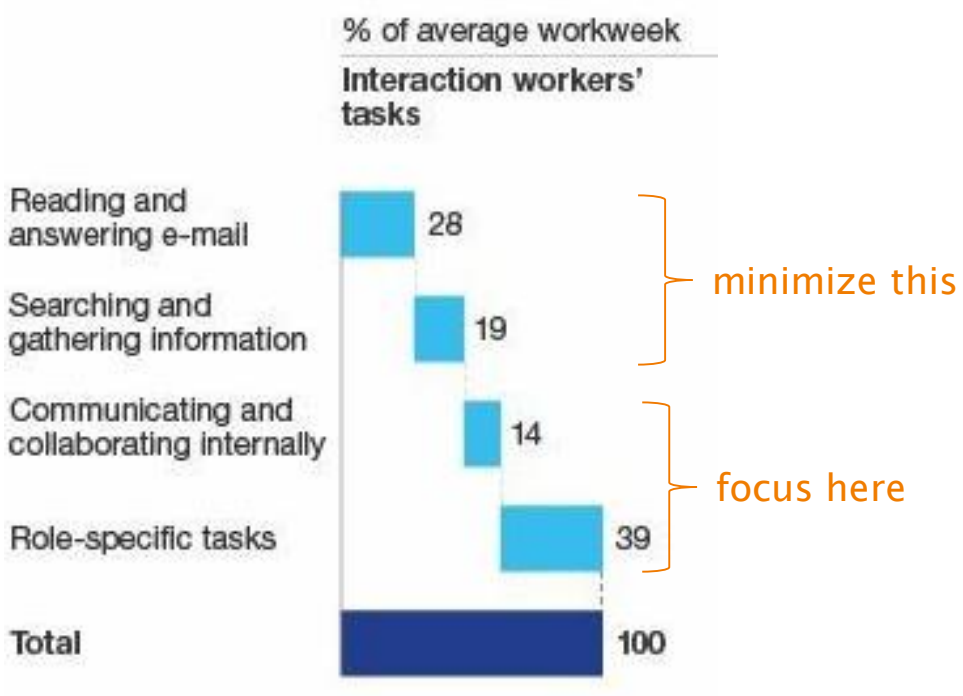
Example Questions:

“Which chromatograms did we make this year for molecule X?”

“Did we already analyze compound X?”

“In which of our labs can I run experiment X?”

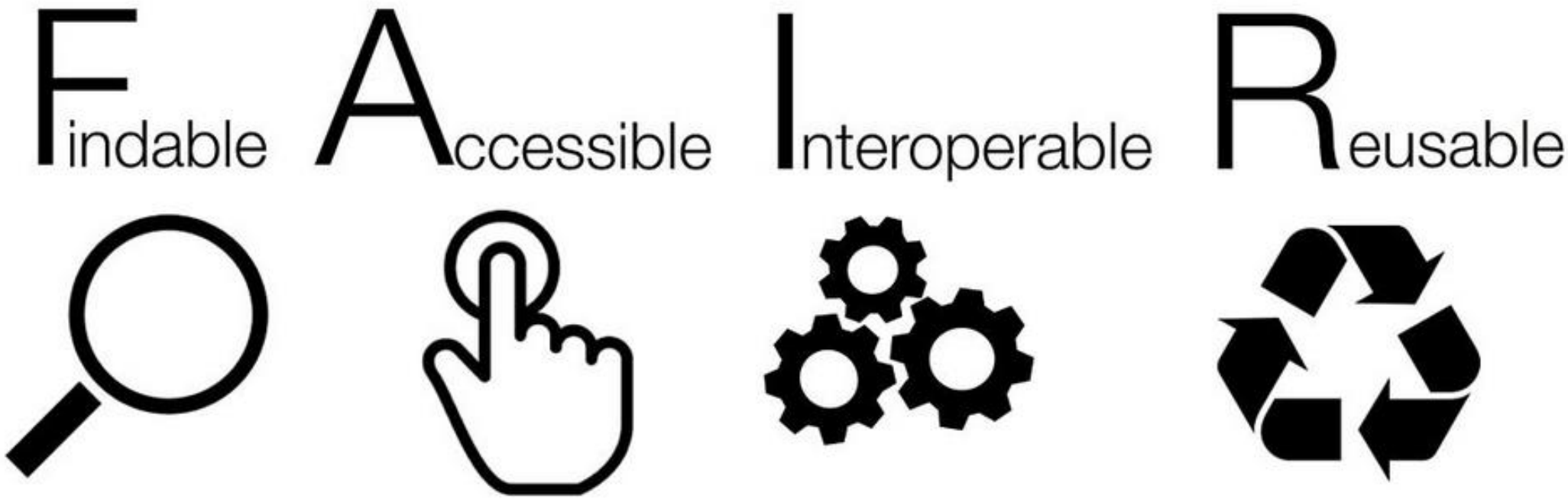
“Can I trust the data that was generated by someone else in my company?”



Source: International Data Corporation (IDC); McKinsey Global Institute analysis.

OSTHUS

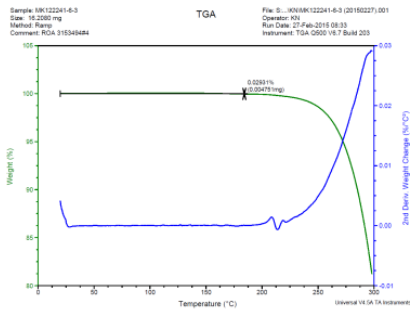
Guiding Principles for Scientific Data Management and Stewardship*



*Source: https://www.nature.com/articles/sdata201618
G20 endorse the FAIR principles: https://www.dtls.nl/2016/09/13/g20-endorse-fair-principles/

OSTHUS

Laboratory Analytical Process



sample

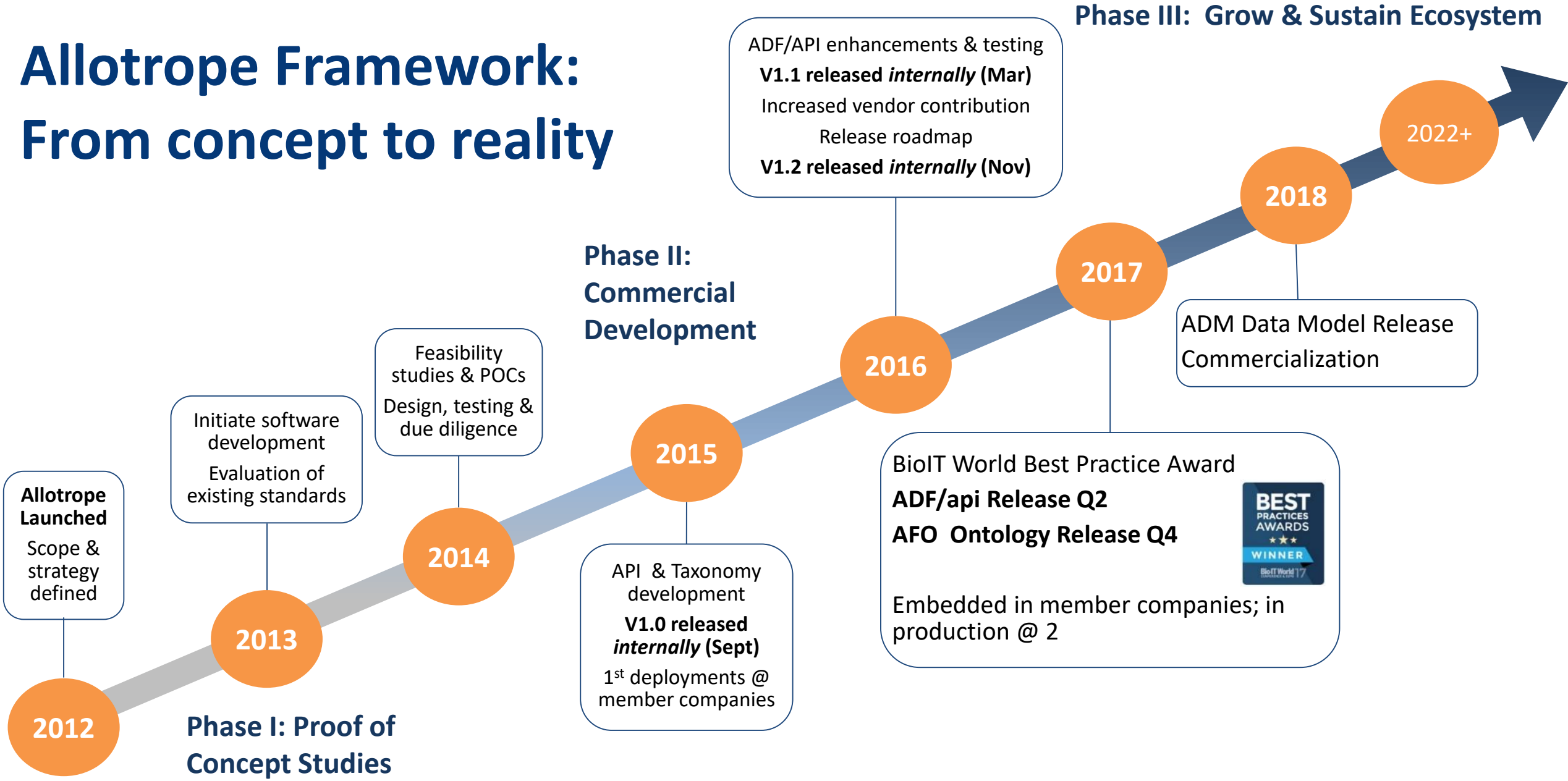
analytical process

data

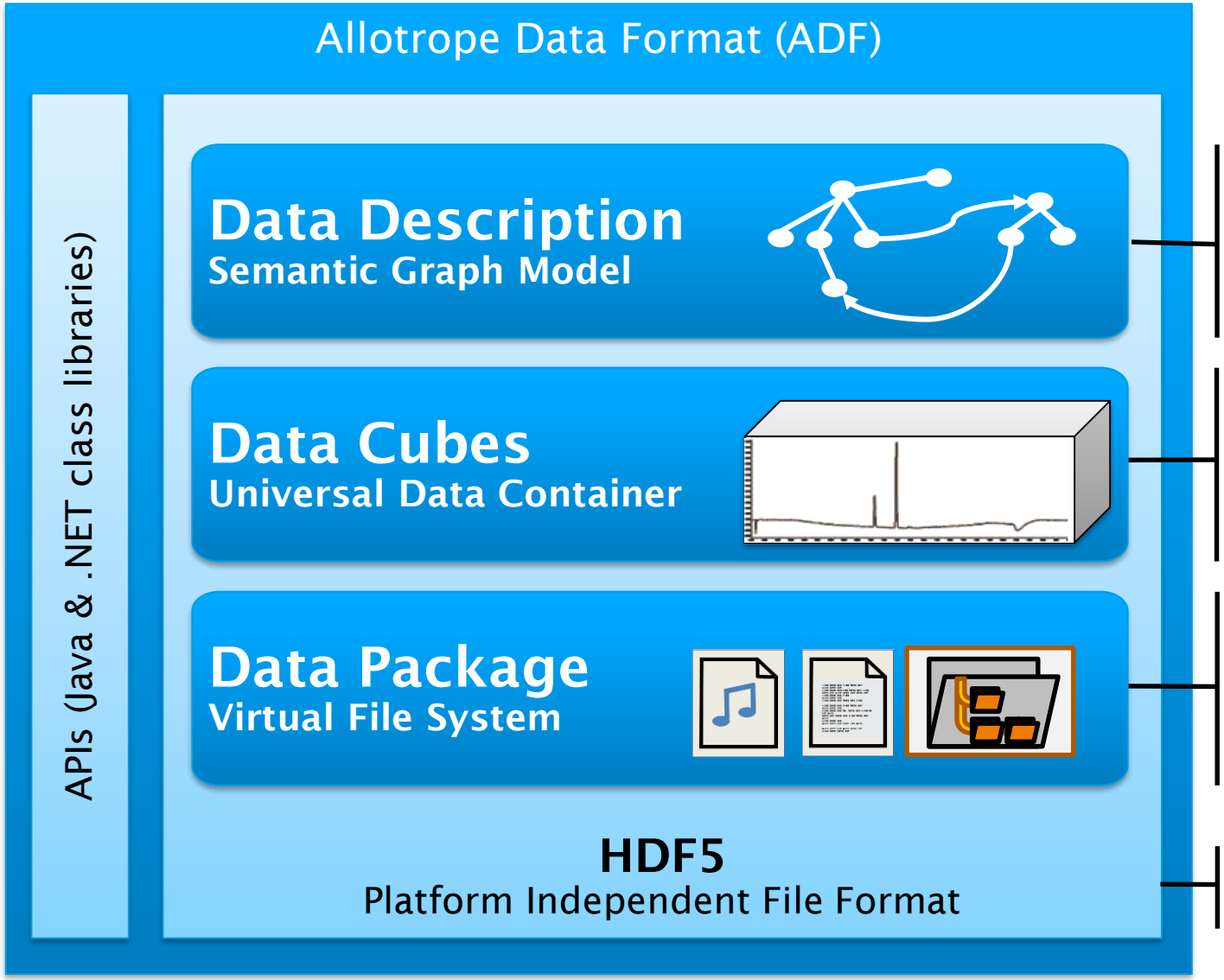
The Allotrope Community Today

Astrix Technology Group • BSSN Software • Elemental Machines • Erasmus MC • Fraunhofer IPA • LabAnswer
Mettler Toledo • NIST • SciBite • Stanford University • University of Illinois at Chicago • University of Southampton

Allotrope Framework: From concept to reality



Allotrope Data Format (ADF): A Universal Data Container



Descriptive metadata about
- Method, instrument, sample, process, result, etc.
- Provenance, audit trail
- Data Cube, Data Package

Analytical data represented by one- or multidimensional arrays of homogeneous data structures.

Analytical data represented by arbitrary formats, incl. native instrument formats, images, pdf, video, etc.

Specifically designed to store and organize large amounts of scientific data.

OSTHUS

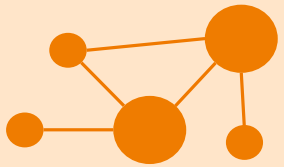
Standardized Data & Metadata

A standardised semantic model for data & metadata.

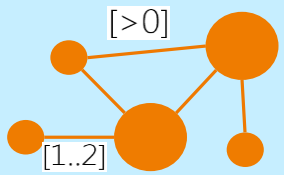
A set of constraints on the semantic model using data shapes.

Released Dec 2017

Ontologies



Data Models



Coming in 2018

Released July 2017

Allotrope Data Format

Data Description
Semantic Model

Data Cubes
Universal Data Container

Data Package
Virtual file system

A high-performance binary data format. Instrument, vendor, platform agnostic.

API

```
ass Directory
  private Directory[] mySub
  private File[] myFiles;

  public Directory()
  {   int numSubs = (int)(M
      mySubs = new Director
      int numFiles = (int)
      myFiles = new File[nu

      for(int i = 0; i < my
          myFiles[i] = new
      for(int i = 0; i < my
          mySubs[i] = new
```

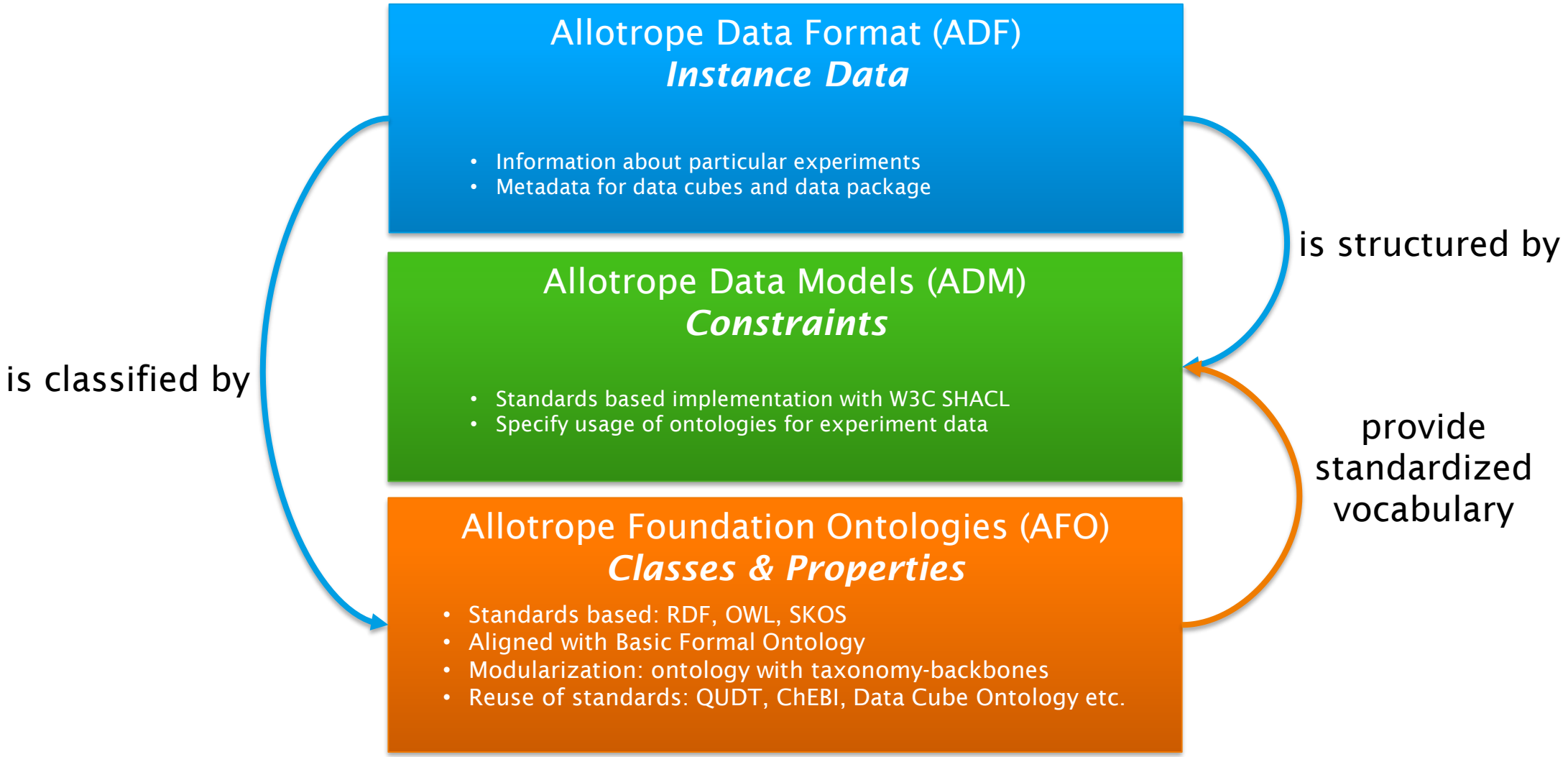
An API to allow consistent creation & reading of ADF files.

ADF Explorer

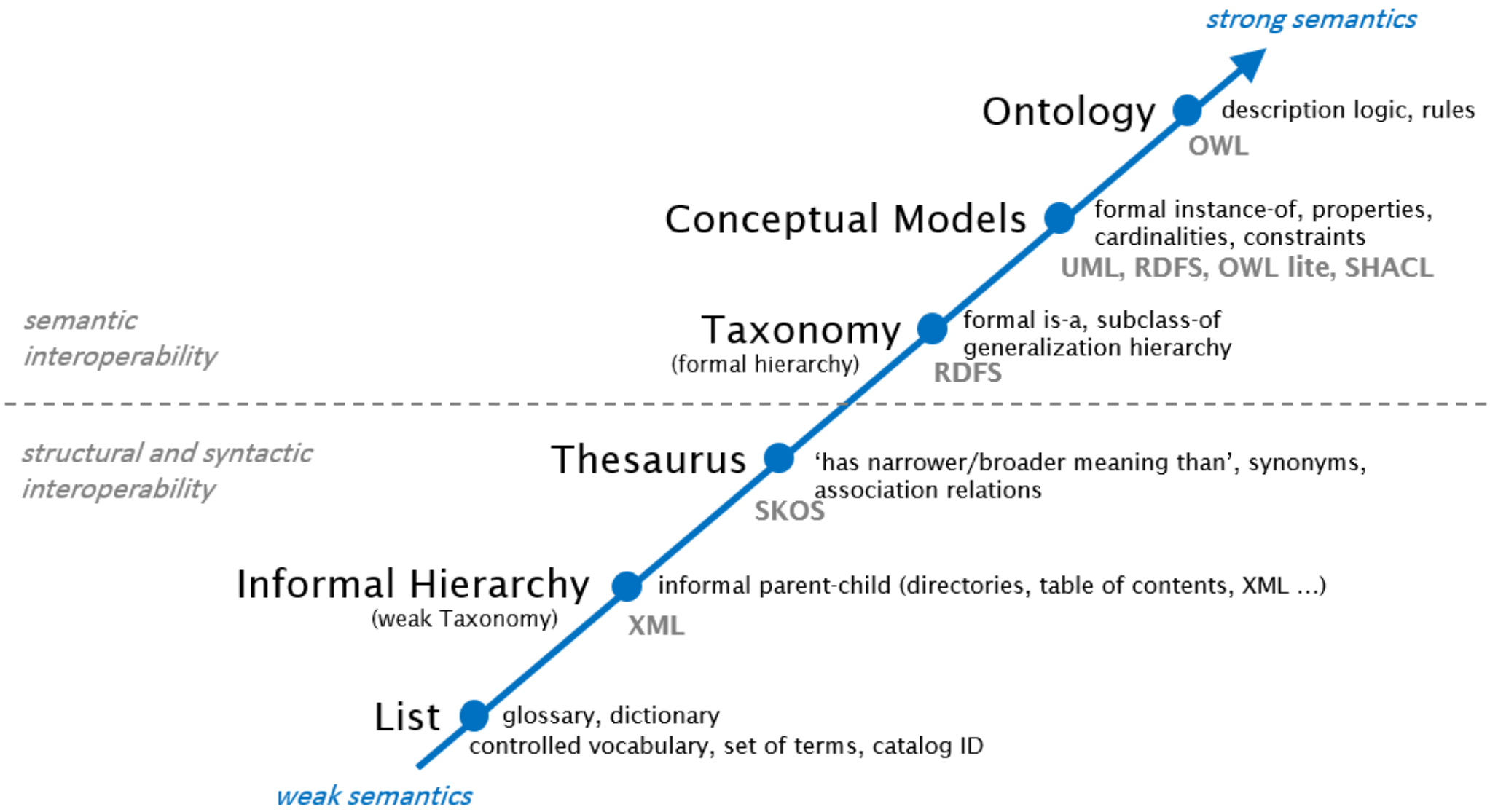
ADF Explorer allows browsing of existing ADF files.

OSTHUS

Semantics: Reference Ontologies, Constraints and Instance Data



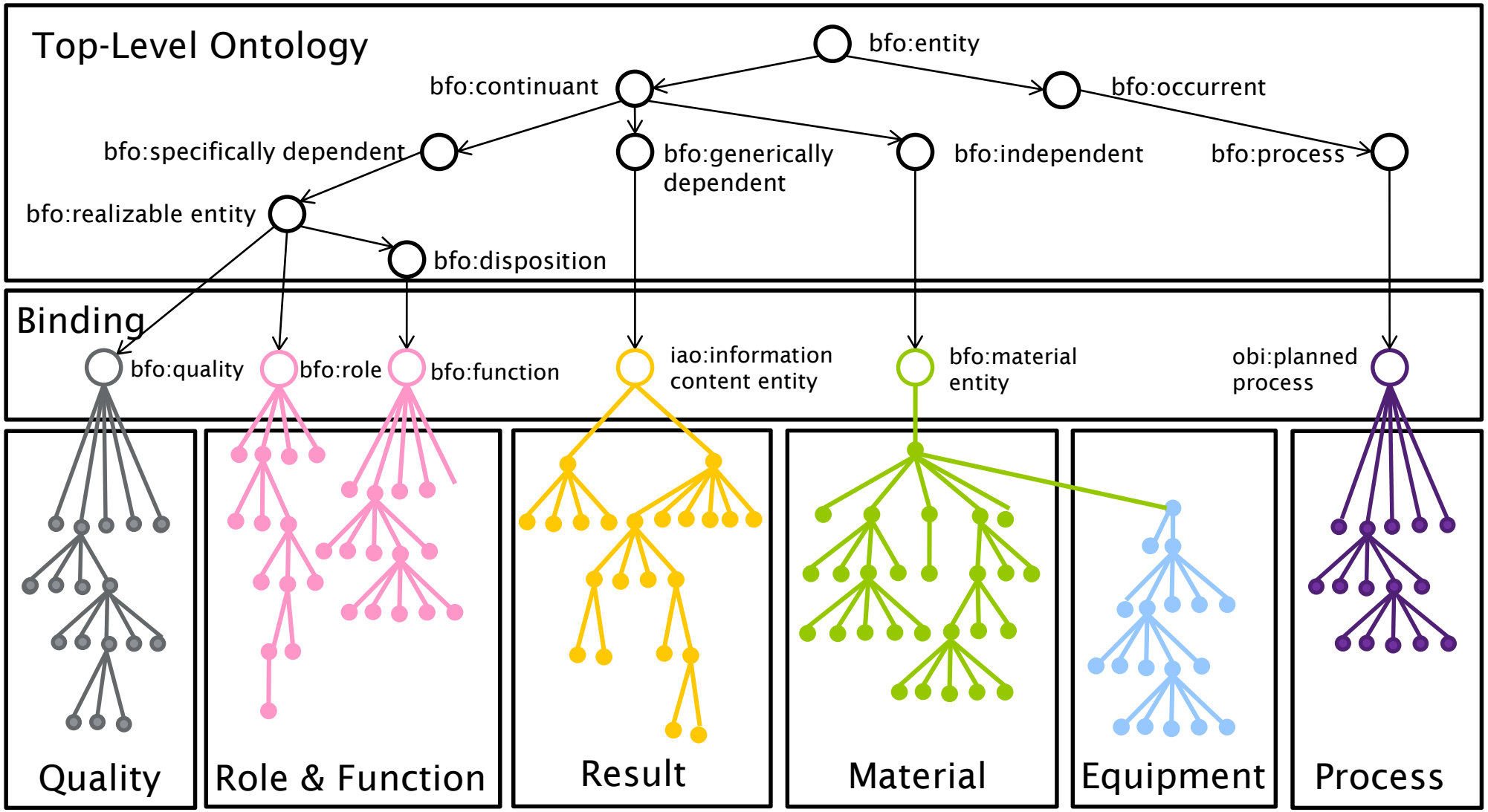
Semantic Spectrum of Knowledge Organization Systems



Sources

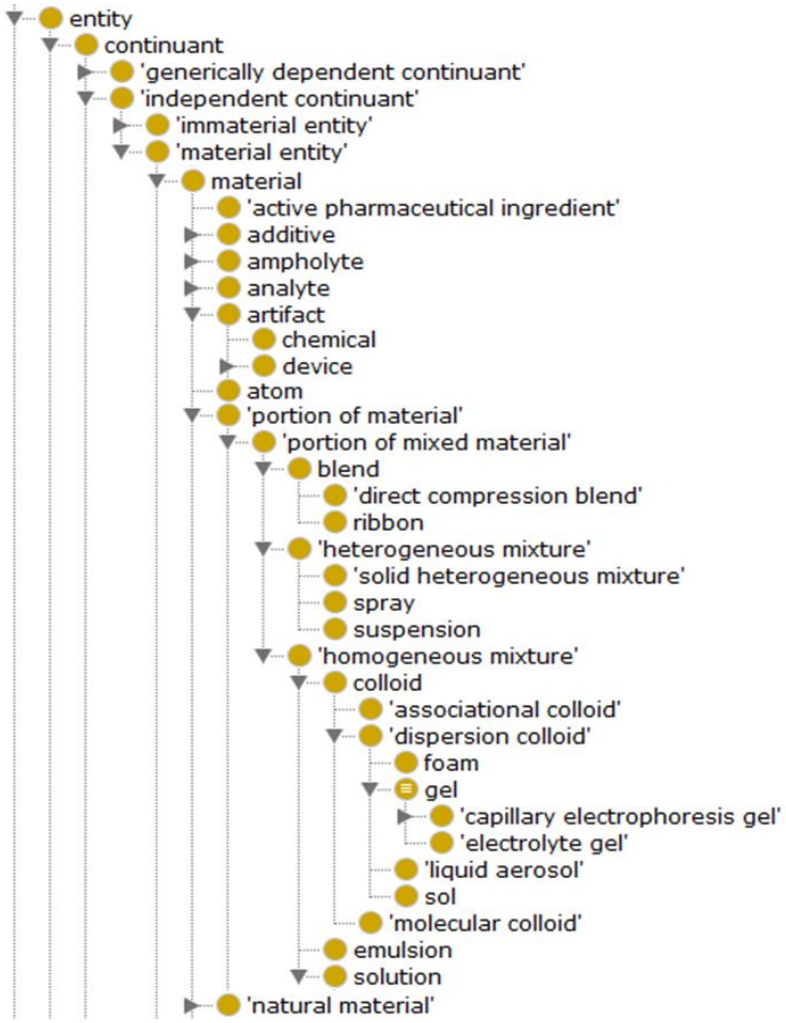
- Deborah L. McGuinness. "Ontologies Come of Age". In Dieter Fensel, Jim Hendler, Henry Lieberman, and Wolfgang Wahlster, editors. Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. MIT Press, 2003.
- Michael Uschold and Michael Gruninger "Ontologies and semantics for seamless connectivity" *SIGMOD Rec.* 33, 4 (December 2004), 58-64. DOI=http://dx.doi.org/10.1145/1041410.1041420
- Leo Obrst "The Ontology Spectrum". Book section in of Roberto Poli, Michael Healy, Achilles Kameas "Theory and Applications of Ontology: Computer Applications". Springer Netherlands, 17 Sep 2010.
- Leo Obrst and Mills Davis "Semantic Wave 2008 Report: Industry Roadmap to Web 3.0 & Multibillion Dollar Market Opportunities". 2008.

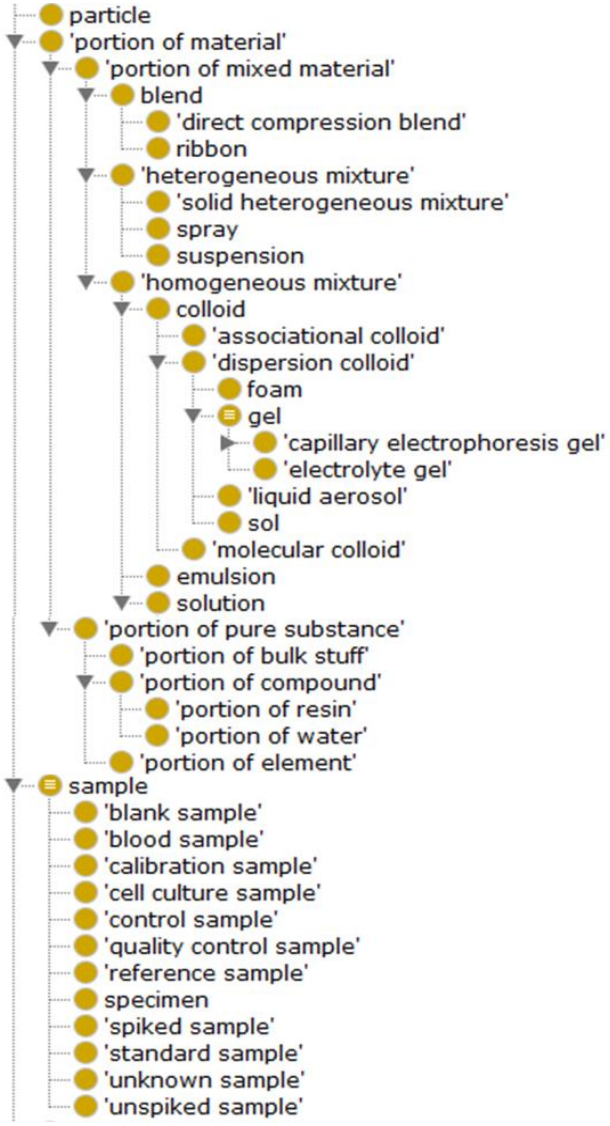
AFO: A Modularized Ontology based on Taxonomy Backbones

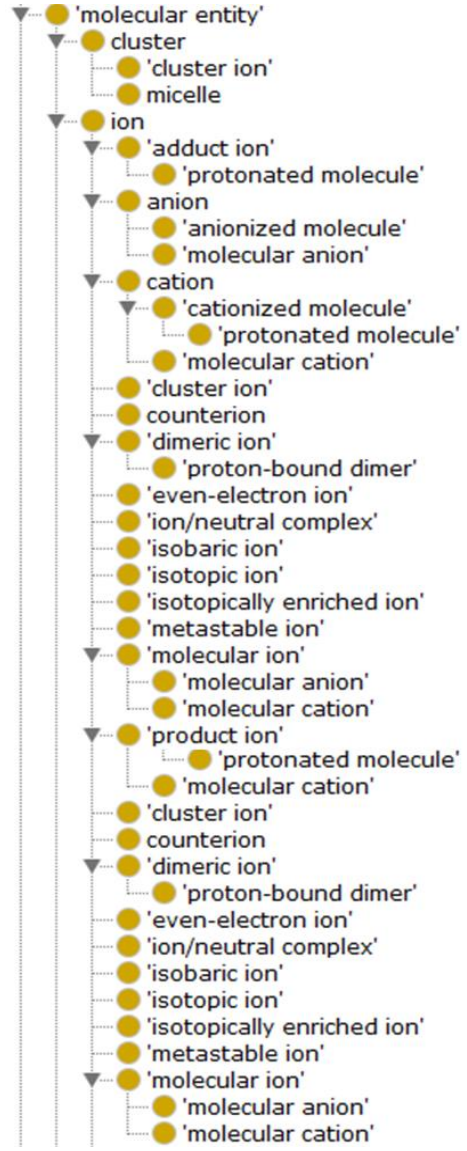


OSTHUS

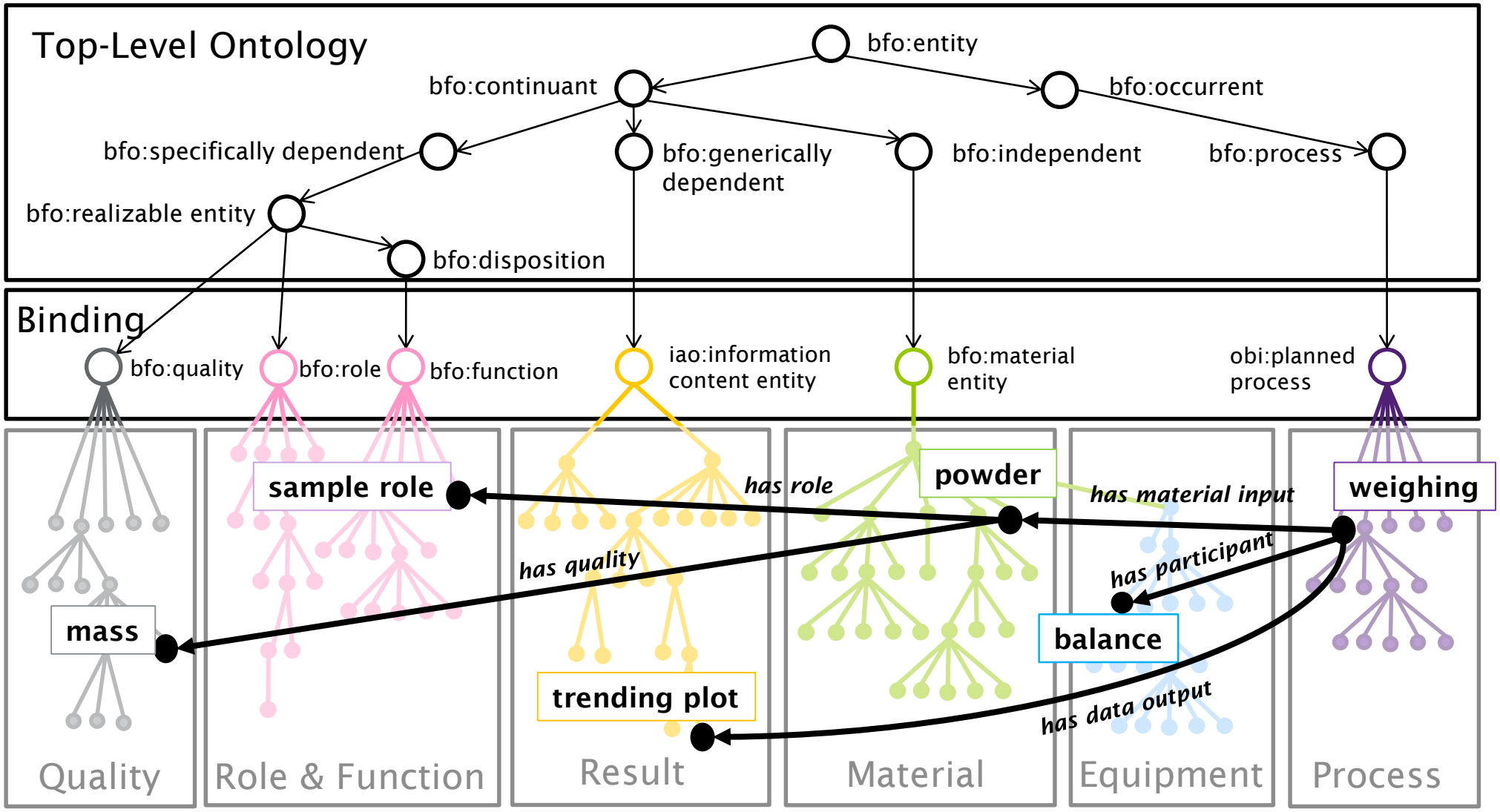
Materials Module (Taxonomy Part)



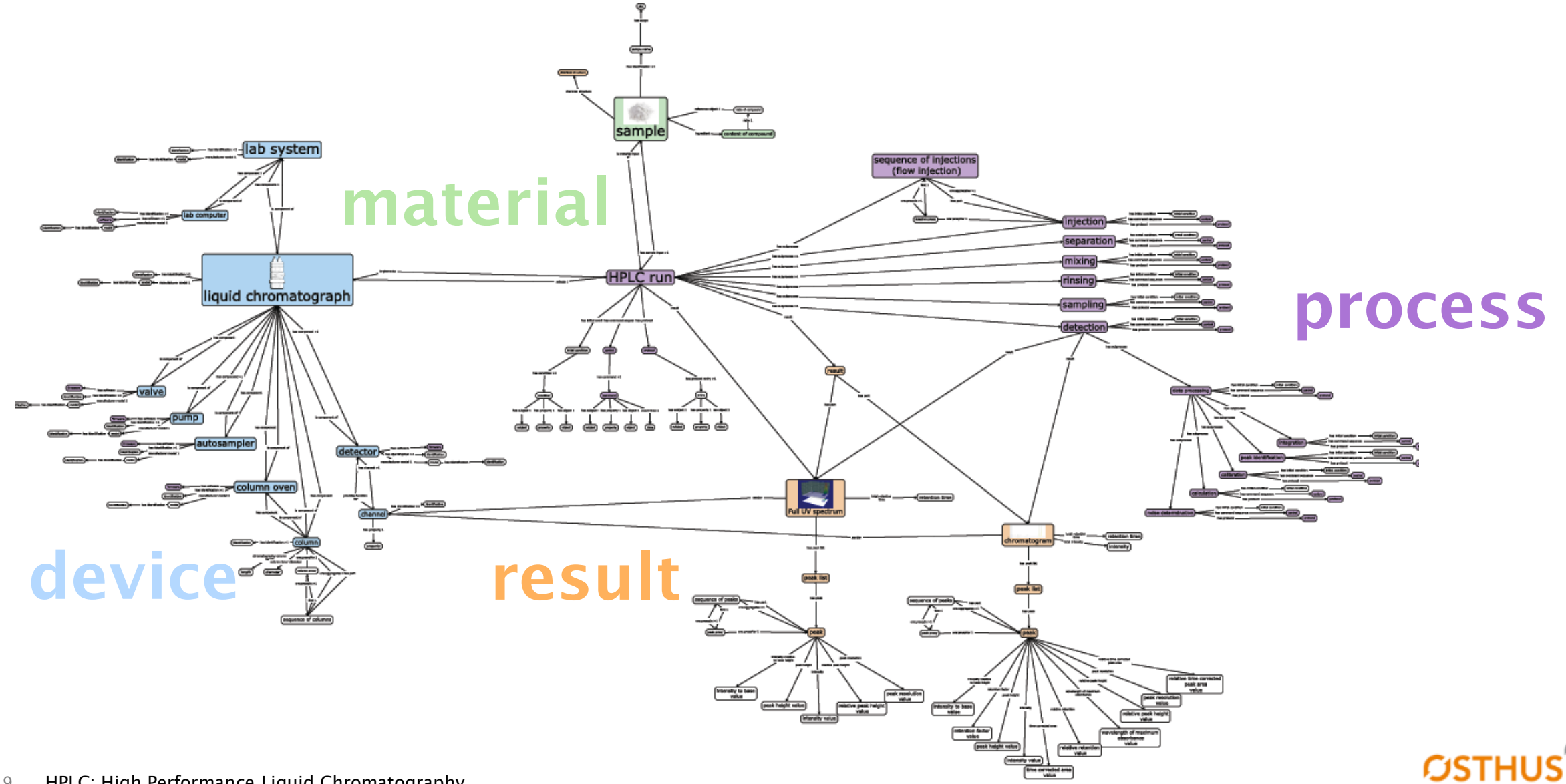




AFO: A Modularized Ontology based on Taxonomy Backbones

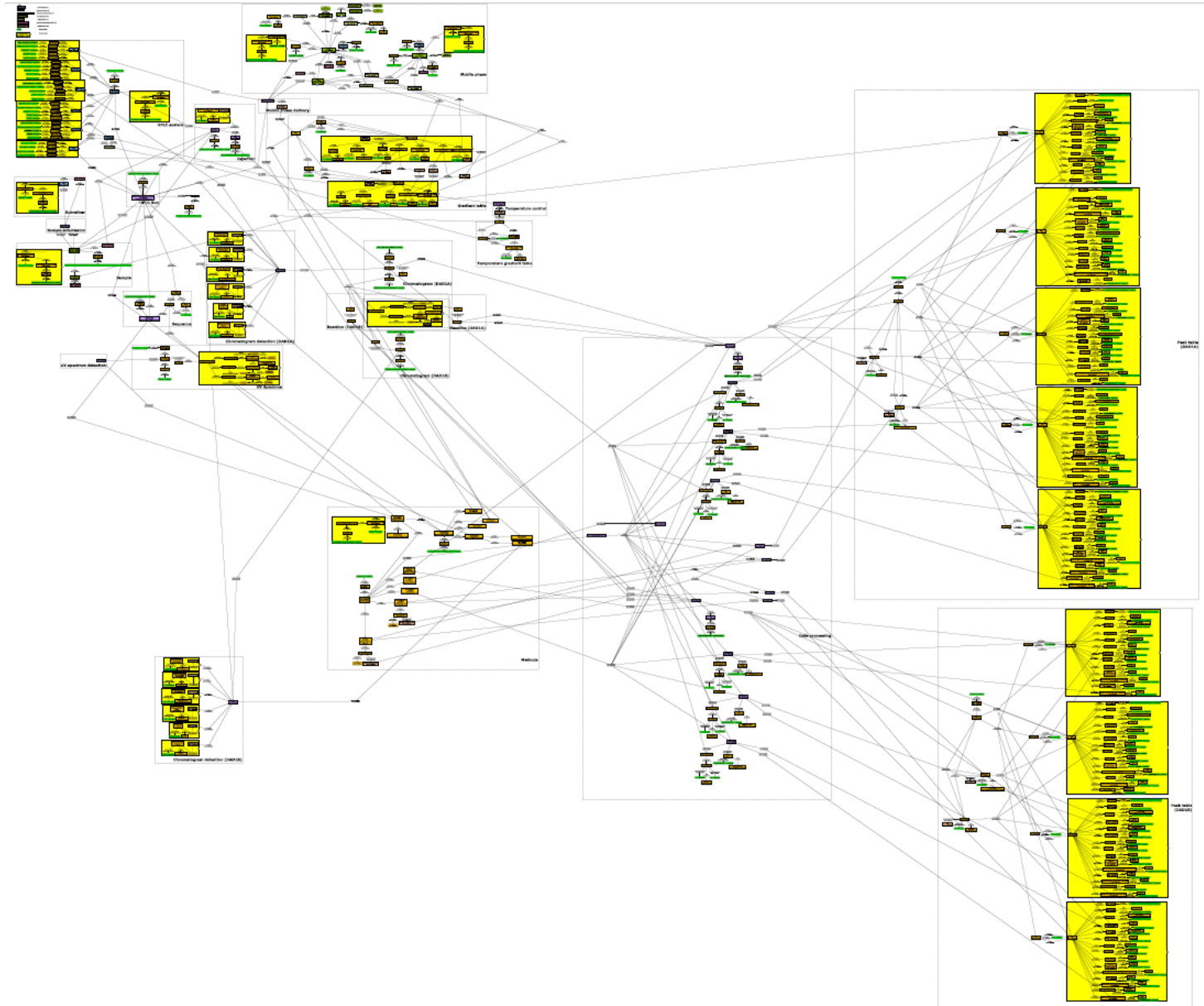


Ontology for HPLC Example



 HPLC: High Performance Liquid Chromatography

Instance Graphs with Repeating Model Patterns



Allotrope Data Model Pattern Catalog

https://allotrope.gitlab.io/adm-patterns/

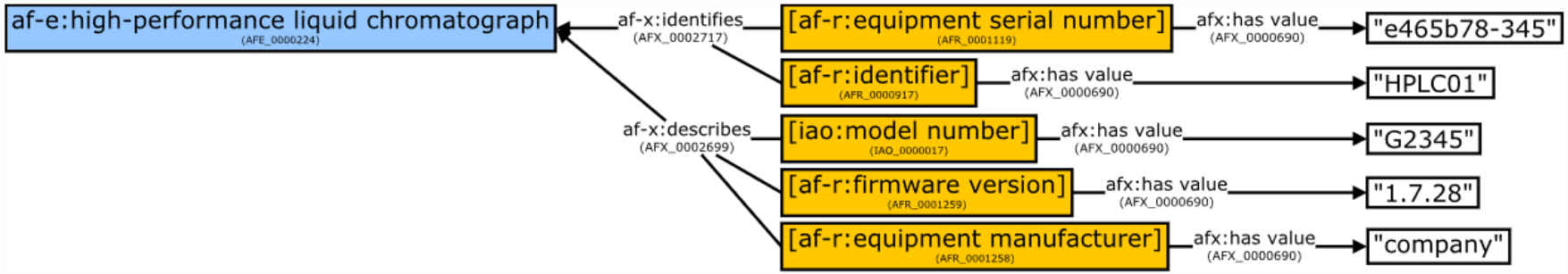
OSTHUS

Allotrope Data Model Pattern Catalog

Allotrope Candidate Recommendation

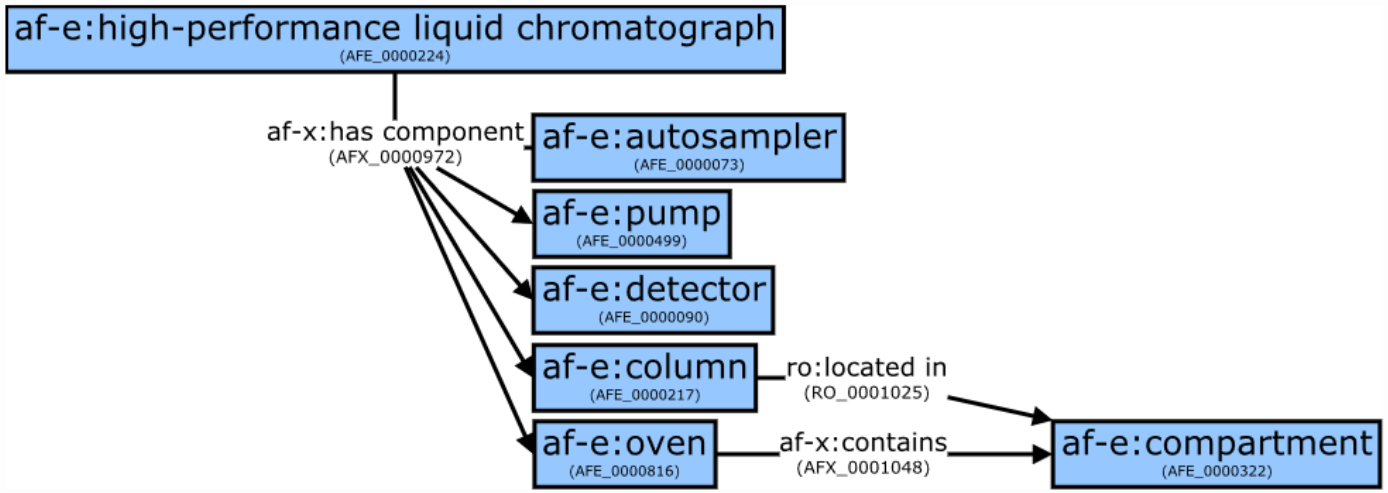
10.1. Device information of an HPLC

Example of information about an high-performance liquid chromatograph.



10.2. Device stack of an HPLC

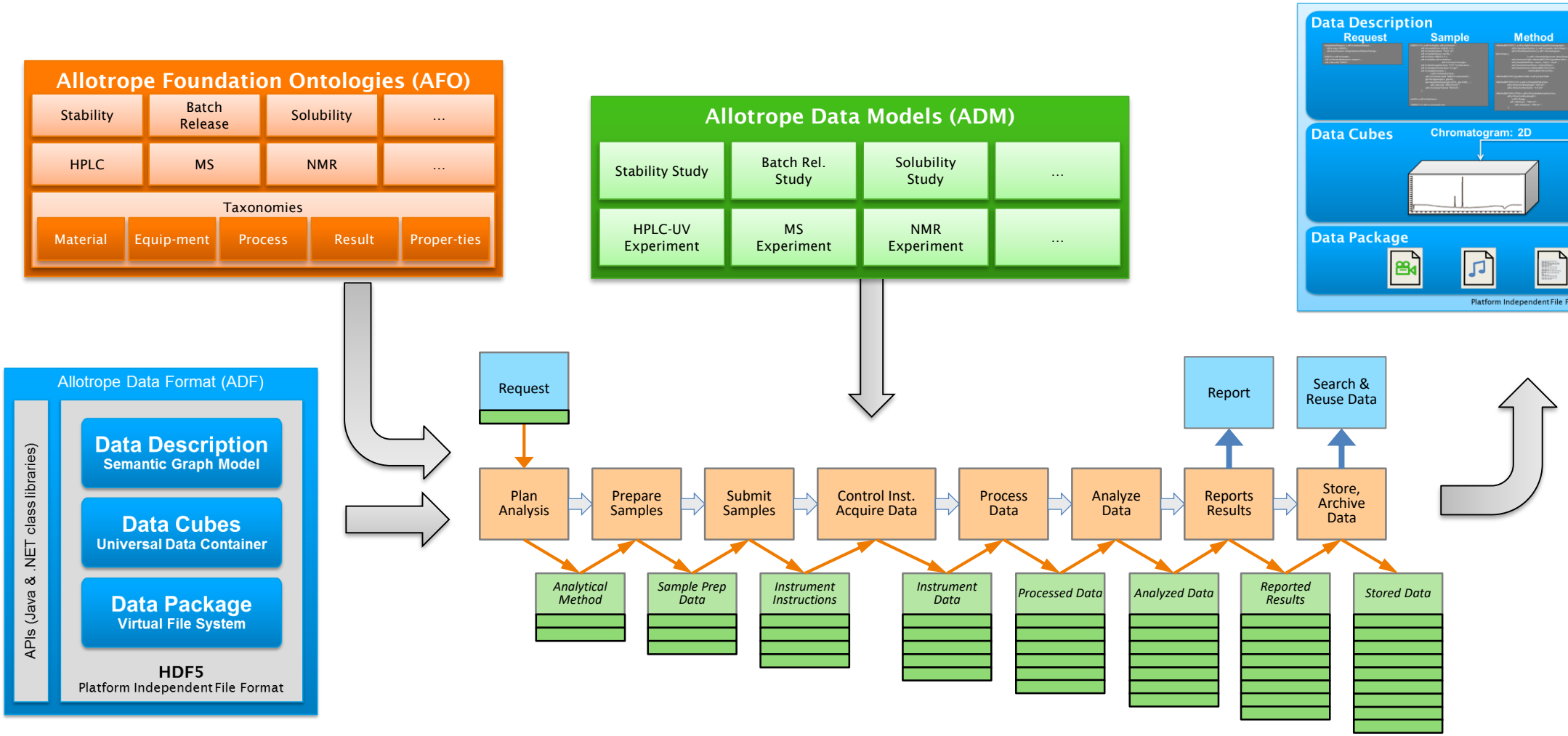
Example of a stack of a high-performance liquid chromatograph. The column is located within a temperature-controlled compartment that is contained in the column oven.



https://allotrope.gitlab.io/adm-patterns/Semantic pattern of device/

OSTHUS

A Foundation for Interoperability & Next Generation Analytics



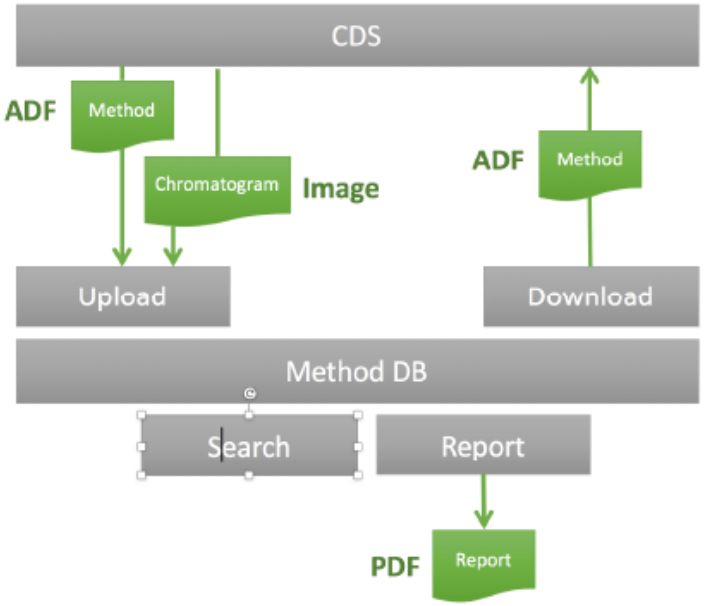
OSTHUS

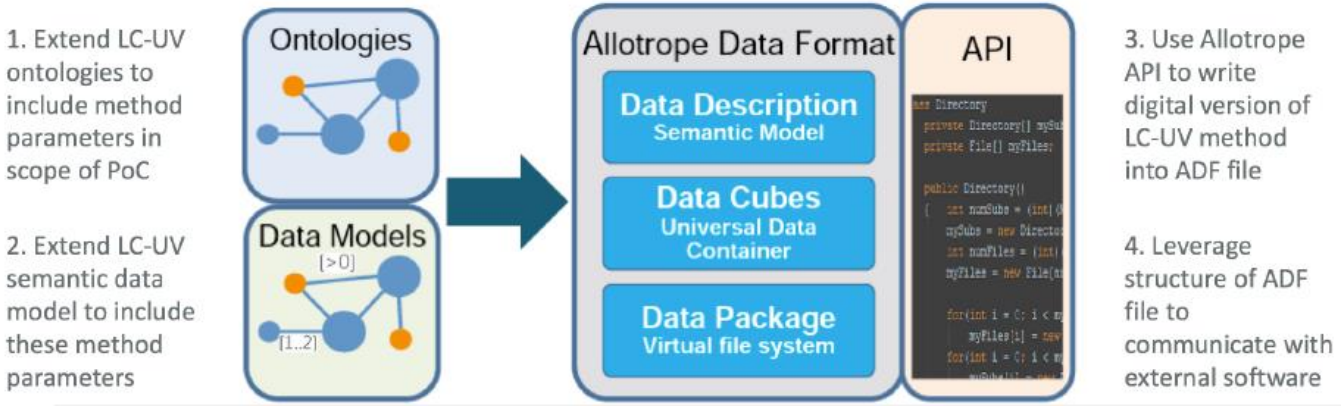
Benefits of Reusing Allotrope Data Format

- Highly reusable, generic data container
  - Flexibility to integrate any custom taxonomy or ontology
  - Can be reused for specification of other standards
- Improves searchability & findability
- Supports interoperability & reusability:
  - Makes data understandable
  - Removes ambiguities
  - Simplifies reproducibility
- Significantly improves information exchange within a community
- Ready for productive use
  - E.g., data integrity, traceability, audit-trail out-of-the-box
- Prepares data for Big Data, Data Science, AI applications

Pistoia Alliance: Methods Database

Method recapitulation across business units within, and external to, pharmaceutical research companies continues to be a challenge. Leveraging **institutional knowledge** to improve method development is also a struggle as this information is neither **shared outside of groups** nor stored in **standard formats** across instrument type. Lastly, making **instruments** and the **discovery process** as a whole, more resilient to local or global **cyber-outages** is becoming a higher priority across IT organizations.





https://www.pistoiaalliance.org/

Archiving Platform

ZONTAL SPACE

- 1st digital archiving platform to enable high data integrity and reuse at enterprise scale.
- Seamless compliance and a greater capacity to innovate.
- Based on open standards and FAIR data principles

ENTERPRISE PLATFORM

Execute cross-functional organizational responsibilities by using a set of integrated and standardized software services that are shared across the enterprise.

Functional needs of individual business units are met by implementing platform-enabled applications.

DATA DISCOVERY AND REUSE

Enhance search capabilities relevant to any global enterprise data application and ensure that non-scientific and scientific data is highly discoverable and enabled for reuse. Finding and extracting layers of standardized metadata during the ingestion process can save an enterprise time and money.

REGULATORY COMPLIANCE

Manage data integrity and access controls across the enterprise. As more data is derived from more sources, centralized data lifecycle management helps companies to both strengthen supervision and enable innovation to take place.

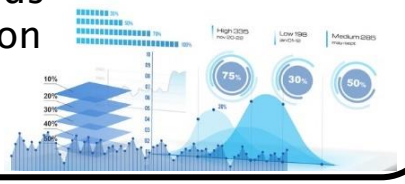
http://www.zontal.io/

OSTHUS

An Information-Centric World Allows to Utilize Data Effectively

Analytics Tools
simulations
statistics

Visualization
dashboards
exploration
search



Reporting
regulatory
internal
external

• • •

Data Science Layer
machine learning, text analytics, NLP, clustering, matching, classification

Lightweight Semantic Integration Layer
data catalogs, reference master data mgt., metadata mgt., semantic indexing, linking, governance, APIs

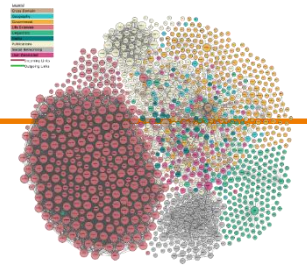
Instrument Data

Operational DBs

Semi-structured Data

Unstructured Documents

Semantic Graph DB



• • •

Linked Open Data & Open APIs

OSTHUS


OSTHUS
Connecting data, people and organizations