V.2.2

From Big Data to Big Analysis

The convergence of Formal Semantics & Data Science in Life Sciences

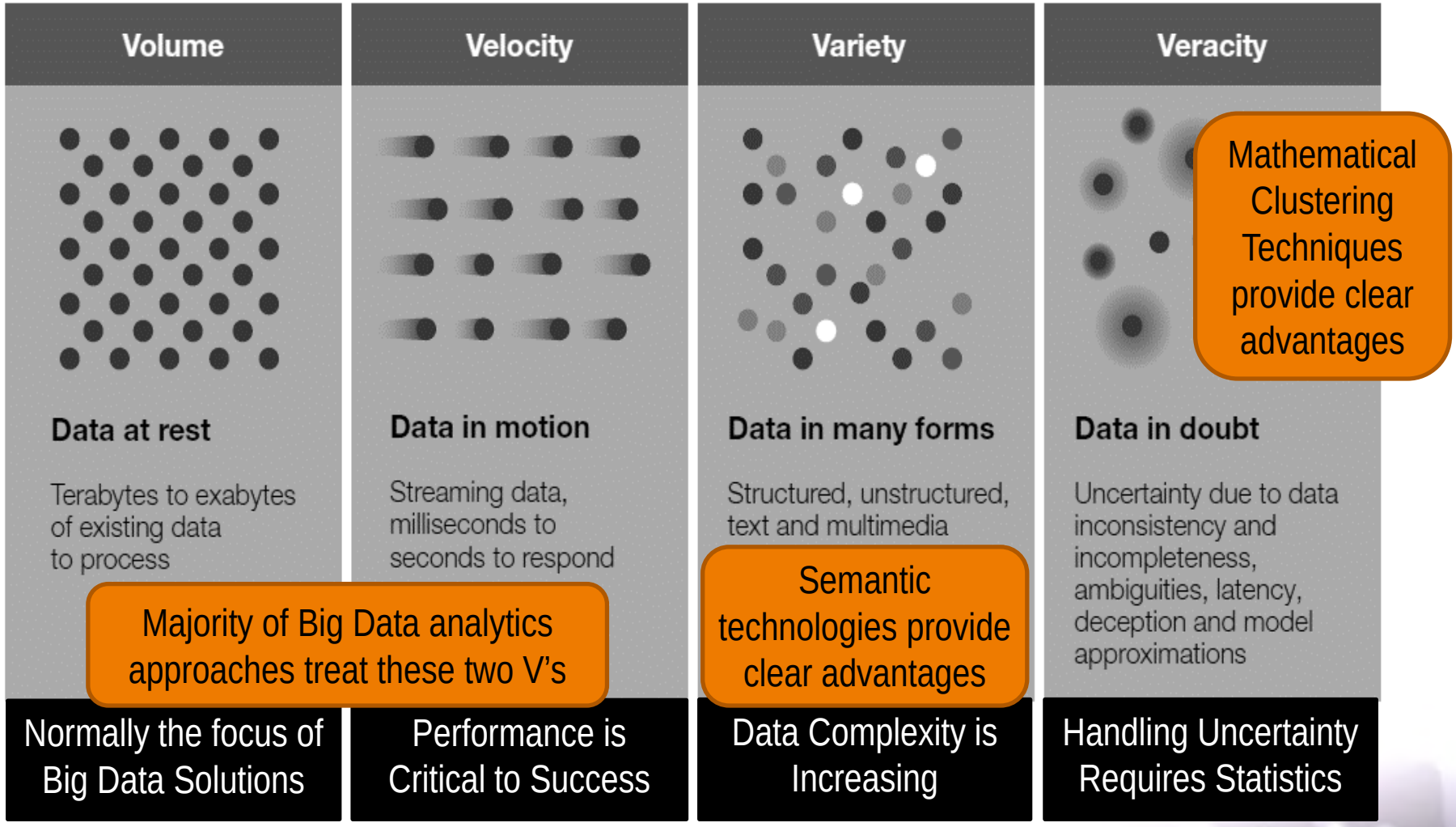
Heiner Oberkampf, PhD

November 7-8th 2017

EMMC Workshop on Interoperability in Materials Modelling, Cambridge

OSTHUS

Understanding the 4V's of Big Data



AT OSTHUS LAB DATA SCIENCE IS

BIG ANALYSIS

STATISTICAL

SEMANTICS

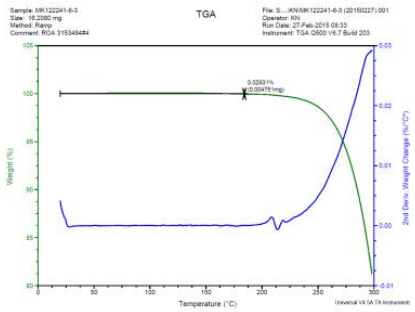
MACHINE LEARNING

REASONING

Laboratory Analytical Process

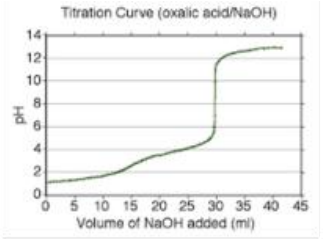
sample

analytical process

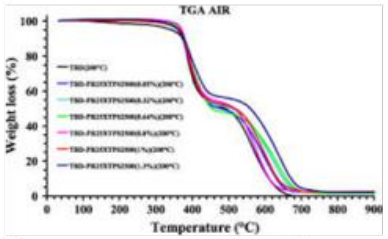


data

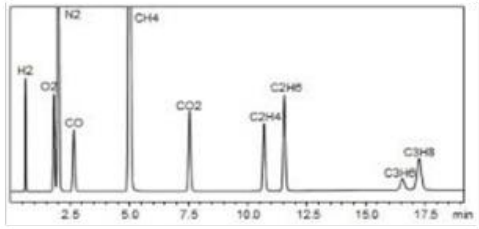
Typical Laboratory Data



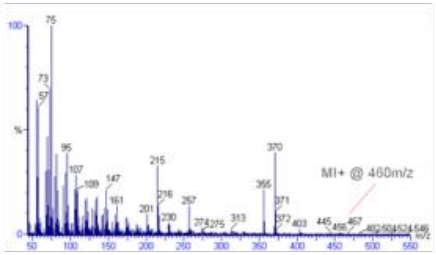
pH



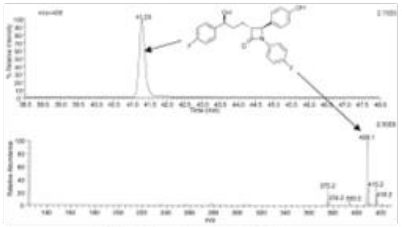
thermogravimetry



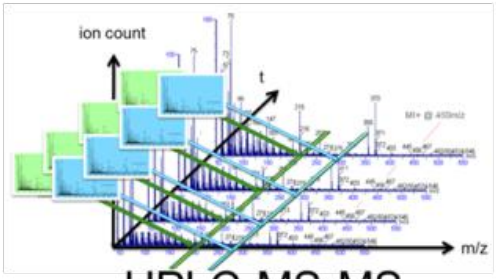
chromatography



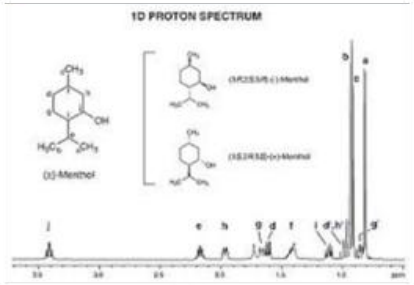
mass spectroscopy



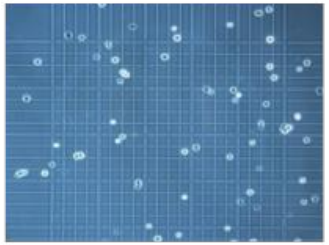
HPLC-MS



HPLC-MS-MS



NMR



cell counter

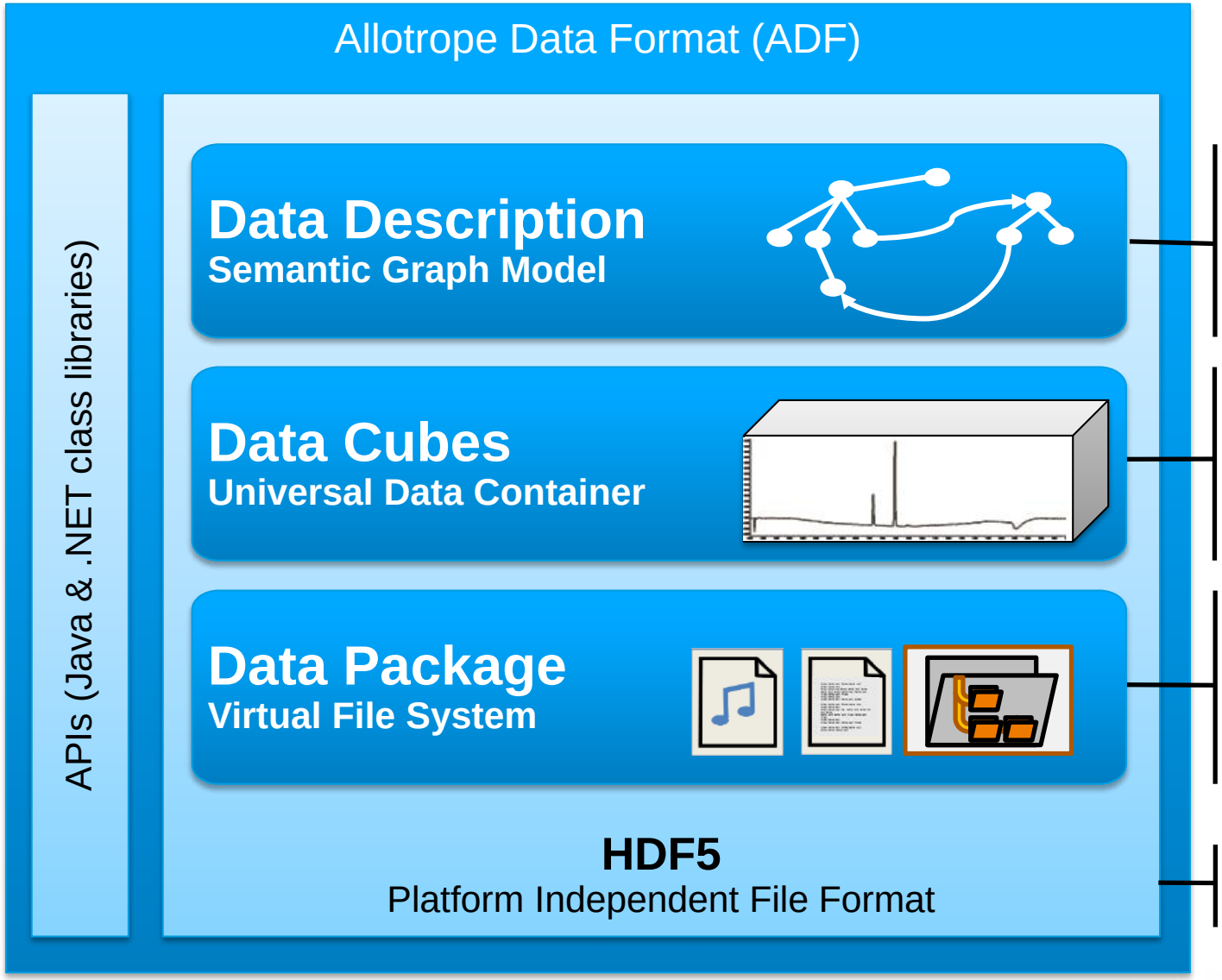
■ ■ ■

Allotrope Structure 2017

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

More information: https://www.allotrope.org/

Allotrope Data Format (ADF)



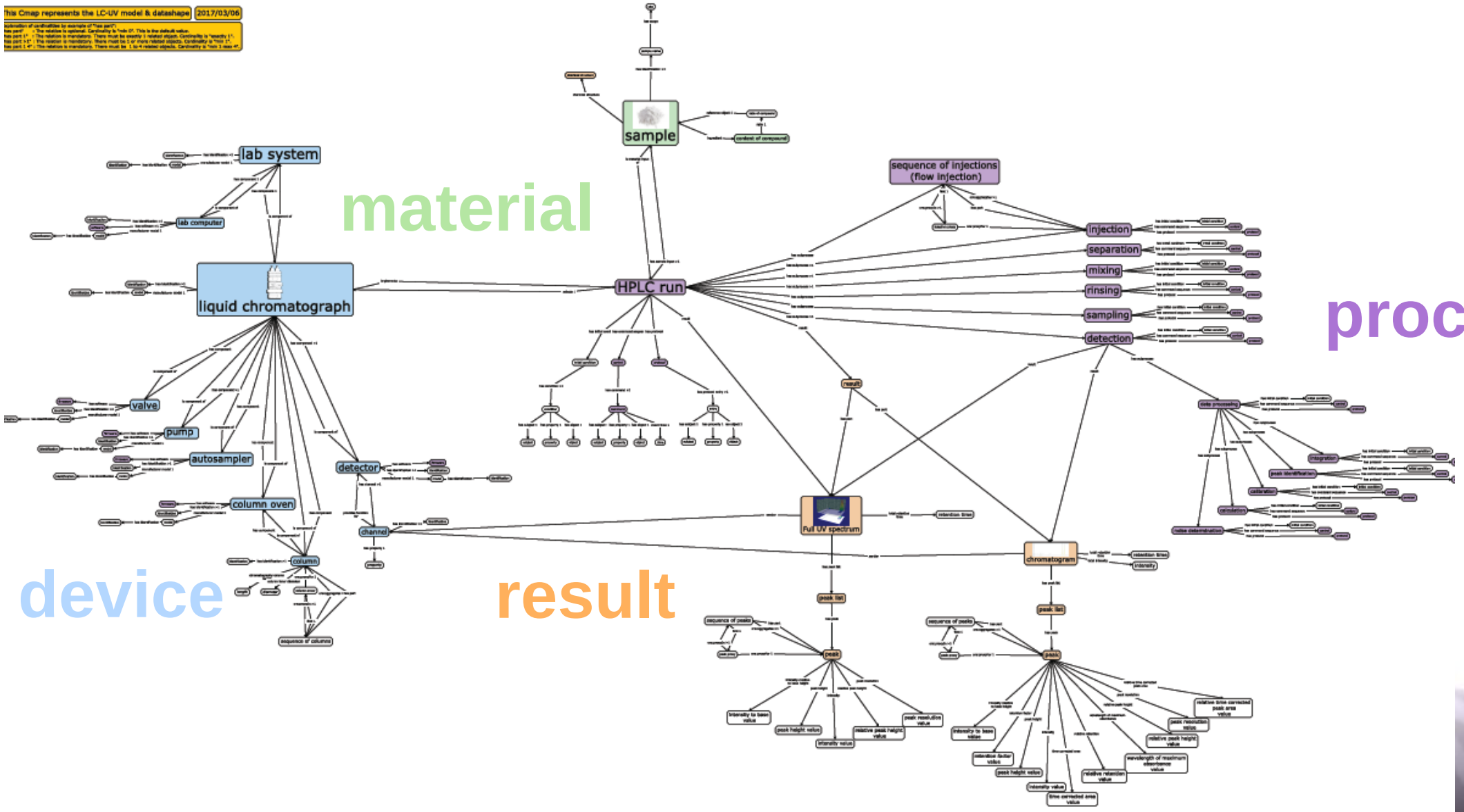
Descriptive metadata about
- Method, instrument, sample, process, result, etc.
- Provenance, audit trail
- Data Cube, Data Package

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

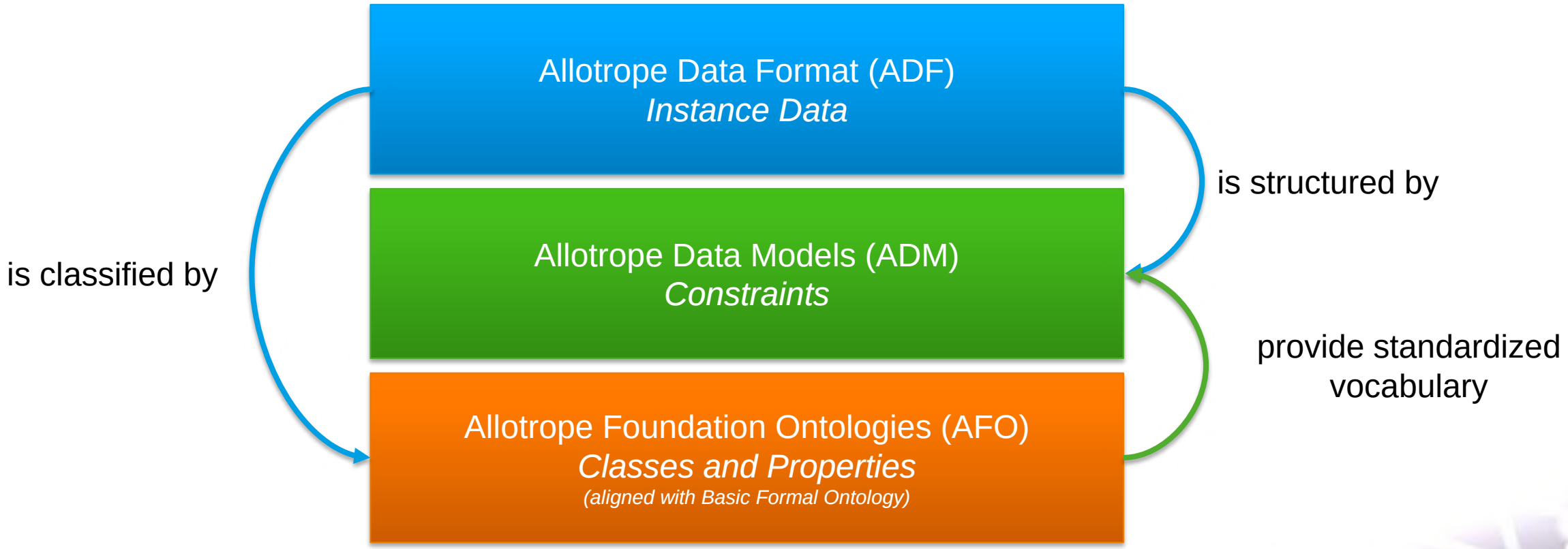
Analytical data represented by arbitrary formats, incl. native instrument formats, images, pdf, video, etc.

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

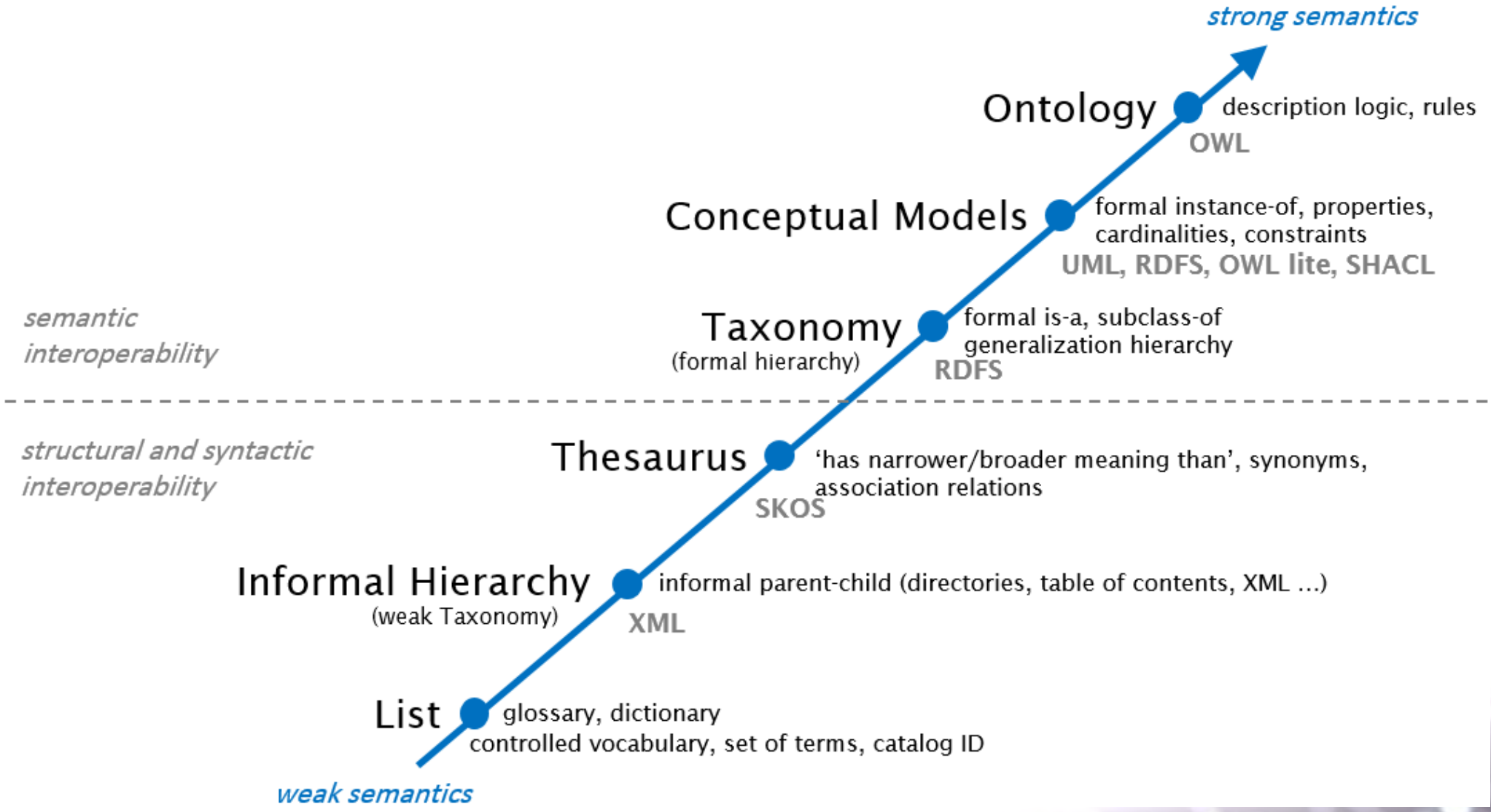
Ontology for HPLC Example



Allotrope Example: Semantics Provides Common Meaning



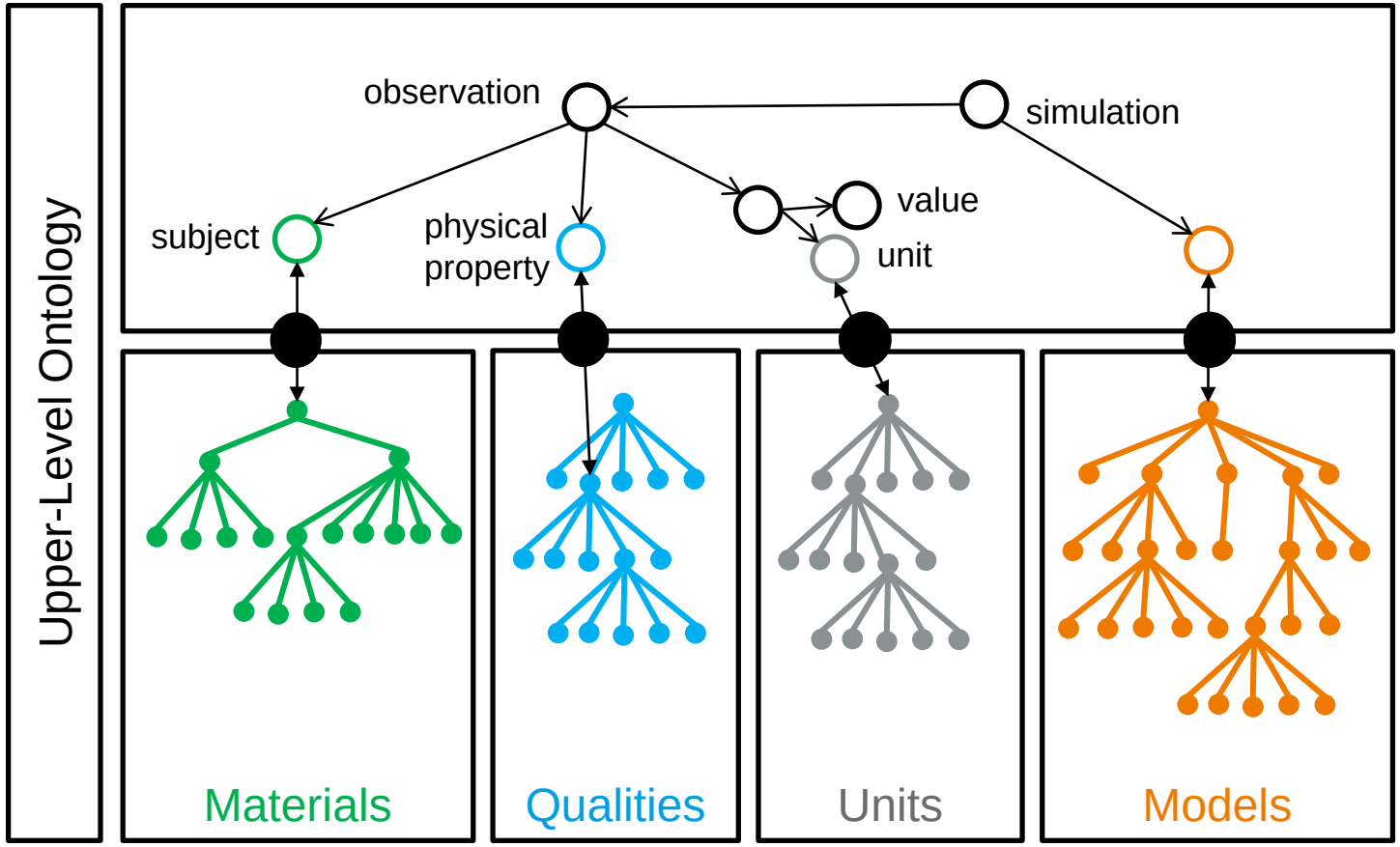
Semantic Spectrum of Knowledge Organization Systems



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.

Application and Reference Ontologies



Application Ontology

- **includes:** Information/Data Model, Schema, Domain Ontology
- **Role:** Defines the important entities and their relationships for a specific application scenario.
- **Realization:** ontology, ER Model, UML etc.

Terminology Binding

- Interface between data model and ref. terminologies

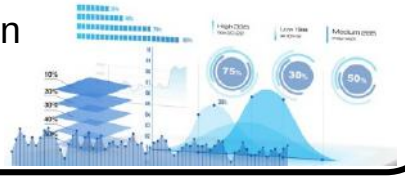
Reference Ontology

- **Also called:** Canonical Reference Ontology, Reference Terminology, Domain Ontology, Foundational Ontology
- **Role:** Standard (structured) vocabulary to be used for placeholder classes of the data model
- **Realization:** list, thesaurus, taxonomy or ontology
- Domain models reusable in many different application scenarios
- **Modules:** Public ref. ontologies plus extension
- Mappings between ref. ontologies

Linked Materials Modelling Data

Visualization

dashboards
exploration
search



Analytics

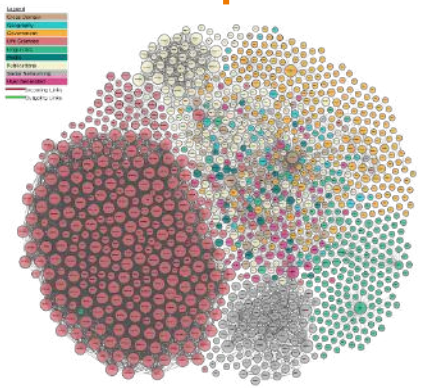
simulations
learning
reasoning

• • •

Lightweight Semantic Integration Layer

Make data **F**indable, **A**ccessible , **I**nteroperable **R**eusable

(APIs, semantic indexing, data annotation, catalogs, metadata and linking)



Linked Open Data
& Open APIs

Semantic
Graph DB
(Knowledge Graph)

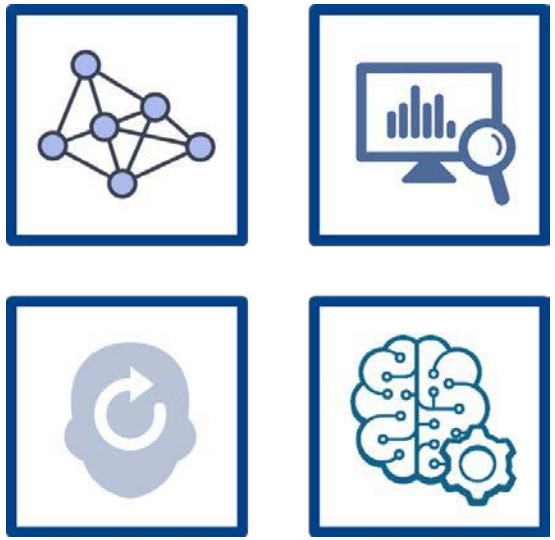
Simulation and
Material Models
Repository

Unstructured
Documents

• • •

The FAIR Guiding Principles for scientific data management and stewardship https://www.nature.com/articles/sdata201618

Towards Big Analysis



1. Think from the end and put use-cases first.
2. Reduce the pain of data sharing and integration by using semantics and FAIR principles.
3. Combine logical and statistical approaches.

OSTHUS

CONNECTING DATA, PEOPLE AND ORGANIZATIONS

Heiner Oberkampf
Consultant at OSTHUS GmbH
+49 (0) 24194314-490
heiner.oberkampf@osthus.com
www.osthus.com