OSTHUS
success with R&D

LIDER workshop, Munich 13th of July 2015

Semantics for Integrated Laboratory Analytical Processes
The Allotrope Perspective

Heiner Oberkampf

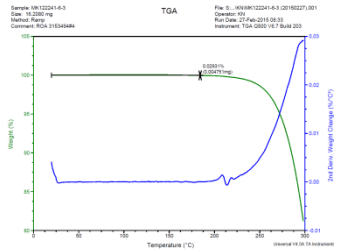
Agenda

- Initial Situation
- Allotrope Foundation
- Approach and IT-Solution
  - Allotrope Data Format
  - Domain Taxonomies
- Use Cases

Laboratory Analytical Processes

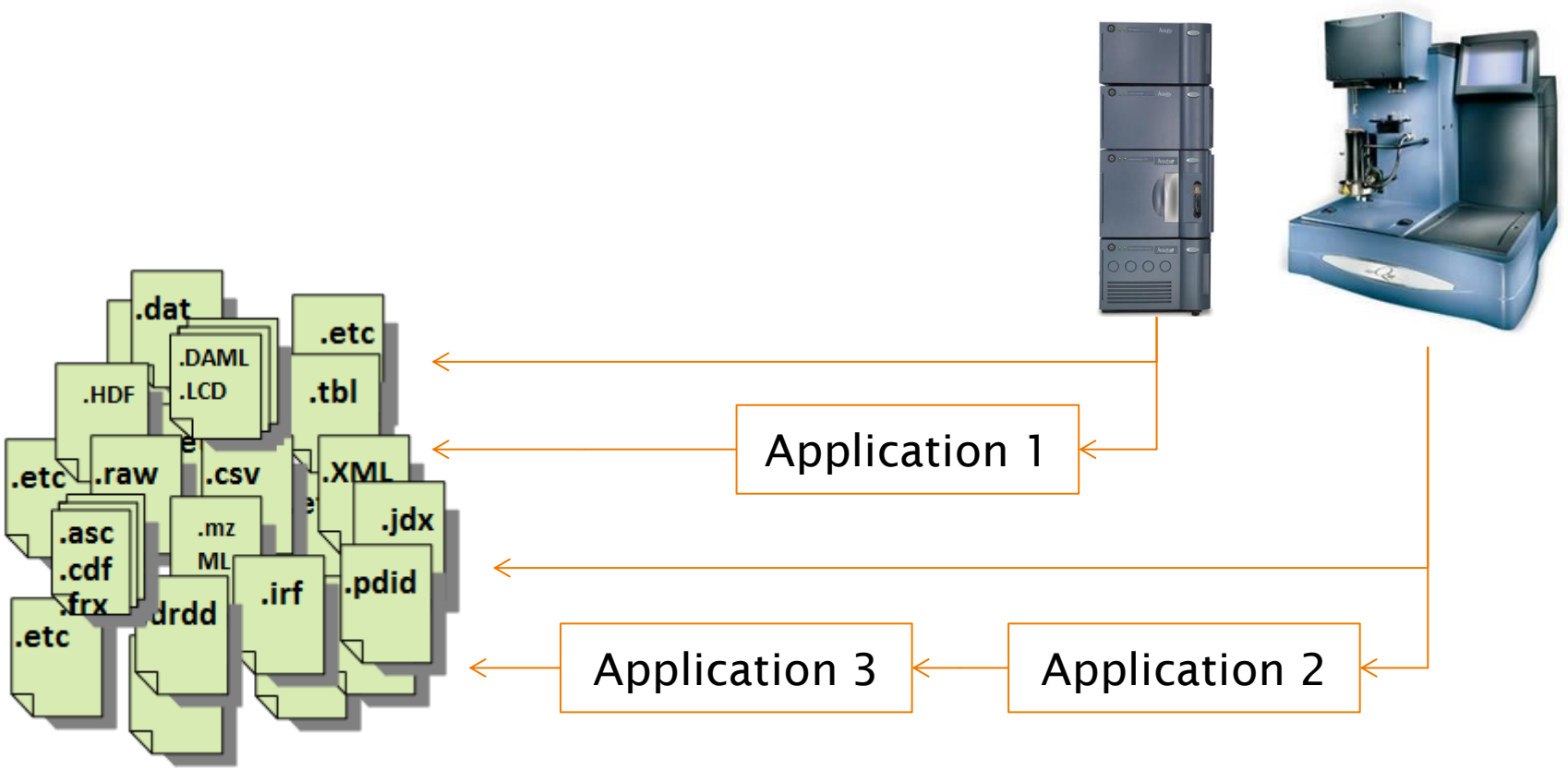
sample

analytical process



data

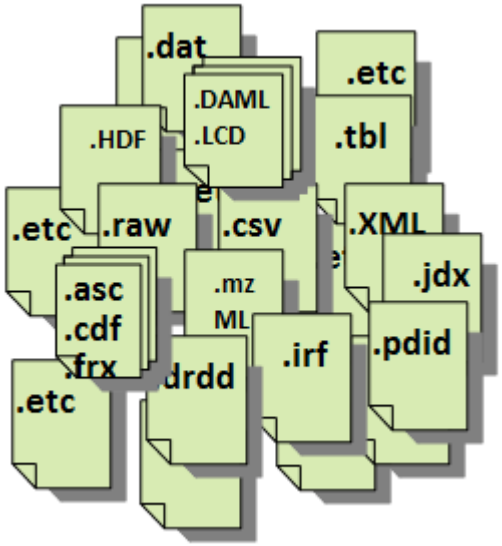
Laboratory Analytical Processes



Common Problems

Instrument & software interoperability is **limited**…at best

It's **hard to mine** a collection of **data** because the **details** and the **context** of the experiment is stored **somewhere else**

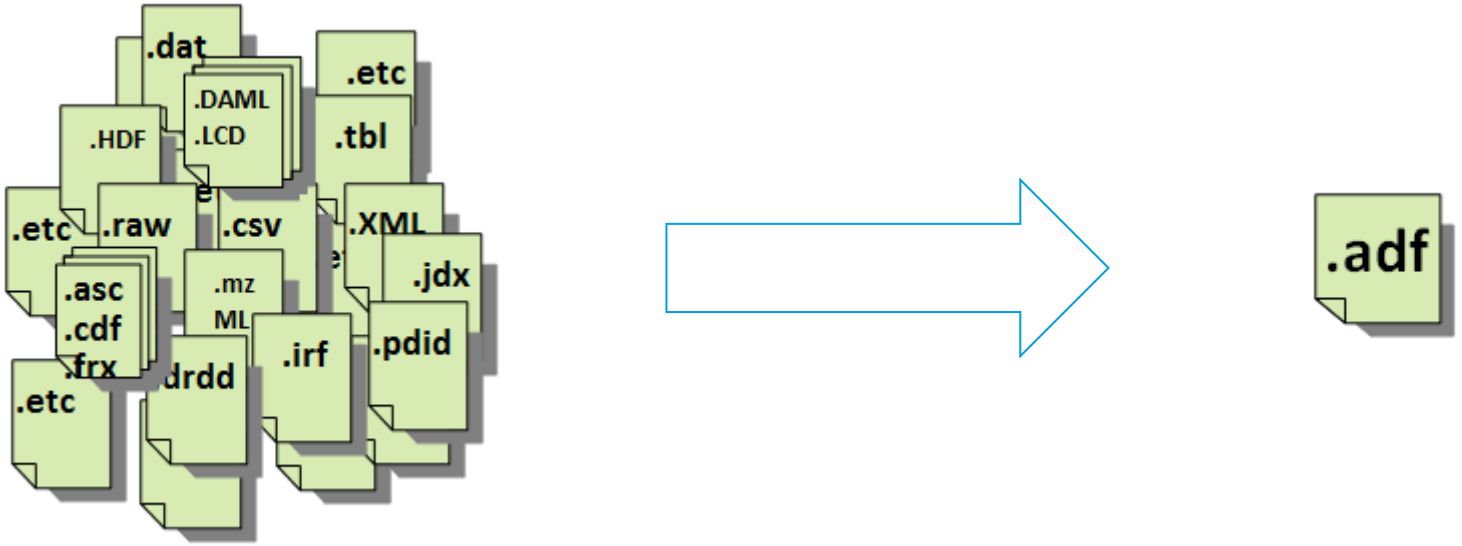


It's **hard to integrate data** from different labs instruments, or online/offline because the **file format is different**

It's hard to find data based on intuitive starting points [e.g. study, project, analyst, technique]

Can't interpret data later because the **context** is **incomplete, inconsistent**, often **free text**

Allotrope Data Format



Allotrope Foundation

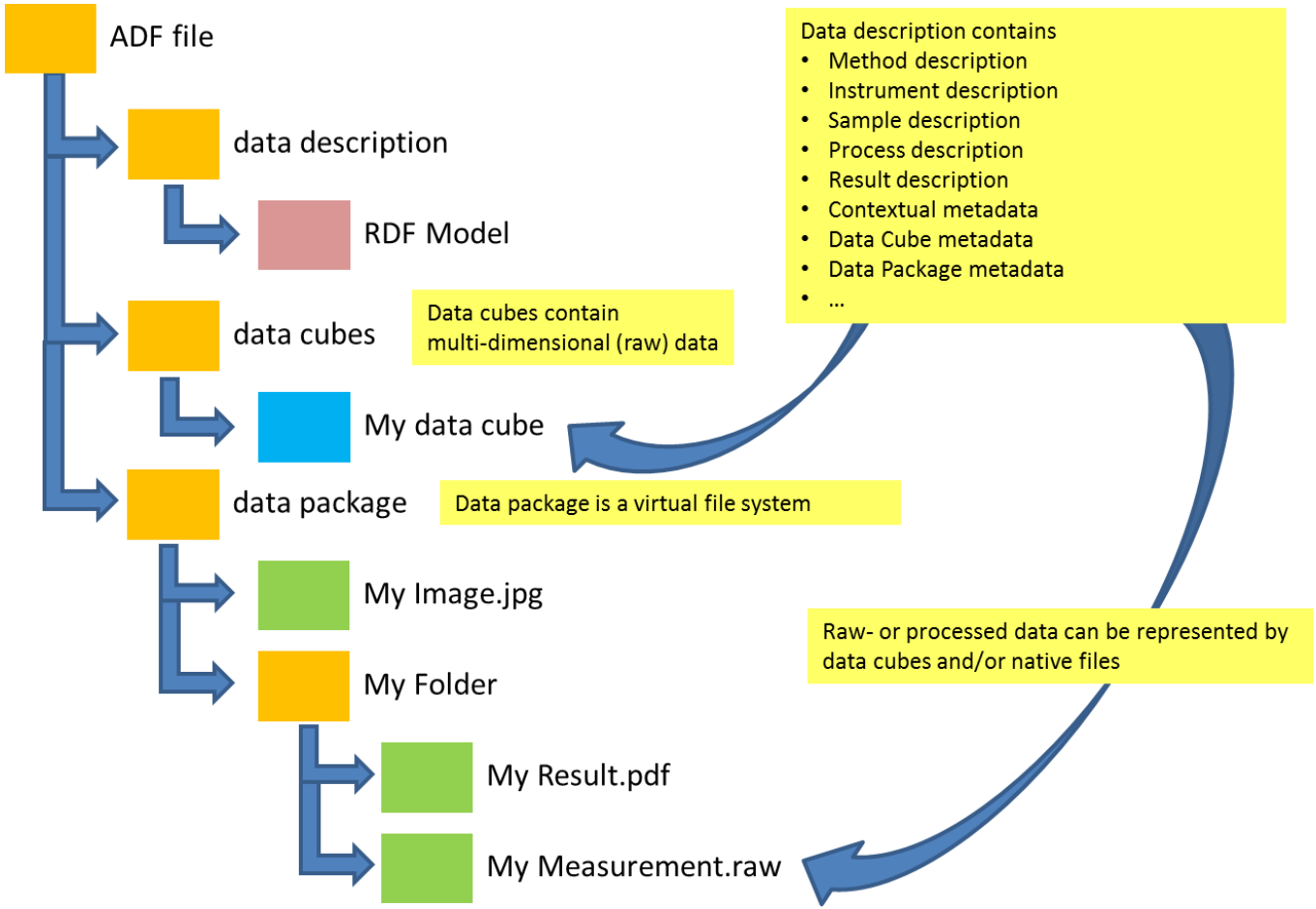
- **Member Companies:** AbbVie, Amgen, Baxter, Bayer, Biogen, Boehringer Ingelheim, Bristol-Myers Squibb, Eli Lilly, Genentech/Roche, GlaxoSmithKline, Merck & Co., Pfizer

- **Secretariat:** Drinker Biddle
  - Project Management
  - Legal & Logistics Support

- **Professional Software Firm:** OSTHUS
  - Framework development
  - Technical leadership

- **Partner Network:** ACD/Labs, Agilent Technologies, BIOVIA, BSSN Software, Erasmus MC, IDBS, Mestrelab Research, Mettler Toledo, Sartorius, Shimadzu, Thermo Scientific, University of Southampton, Waters

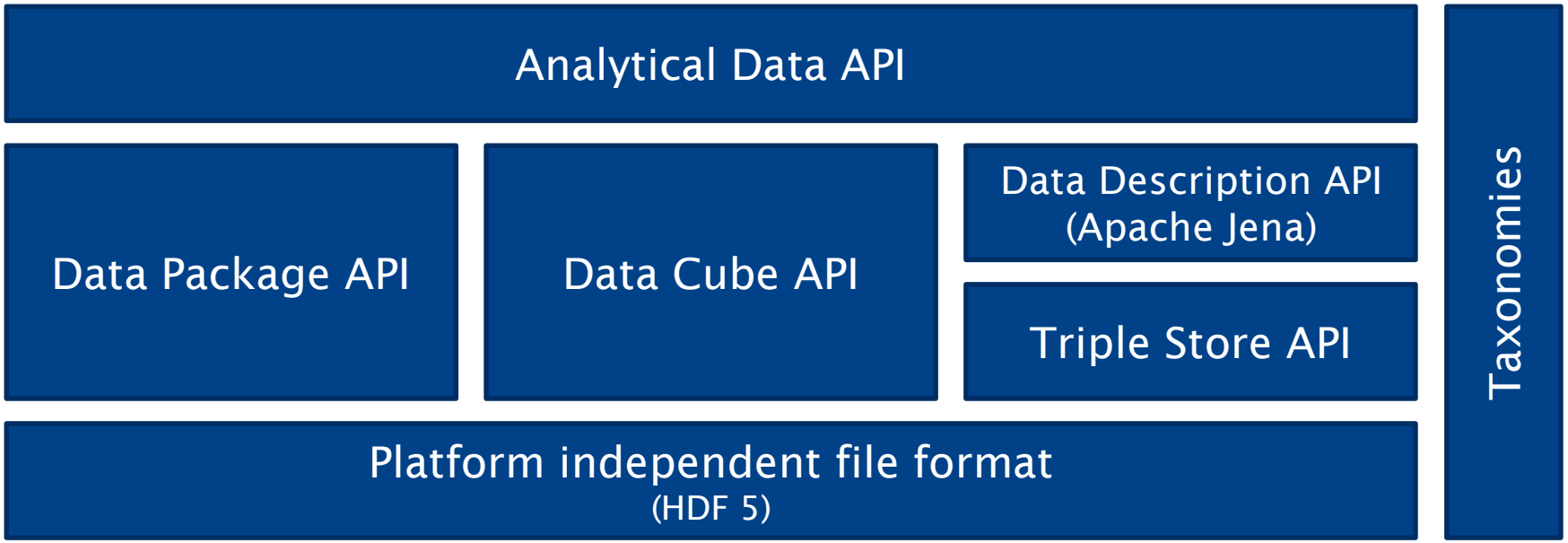
Allotrope Data Format (ADF)



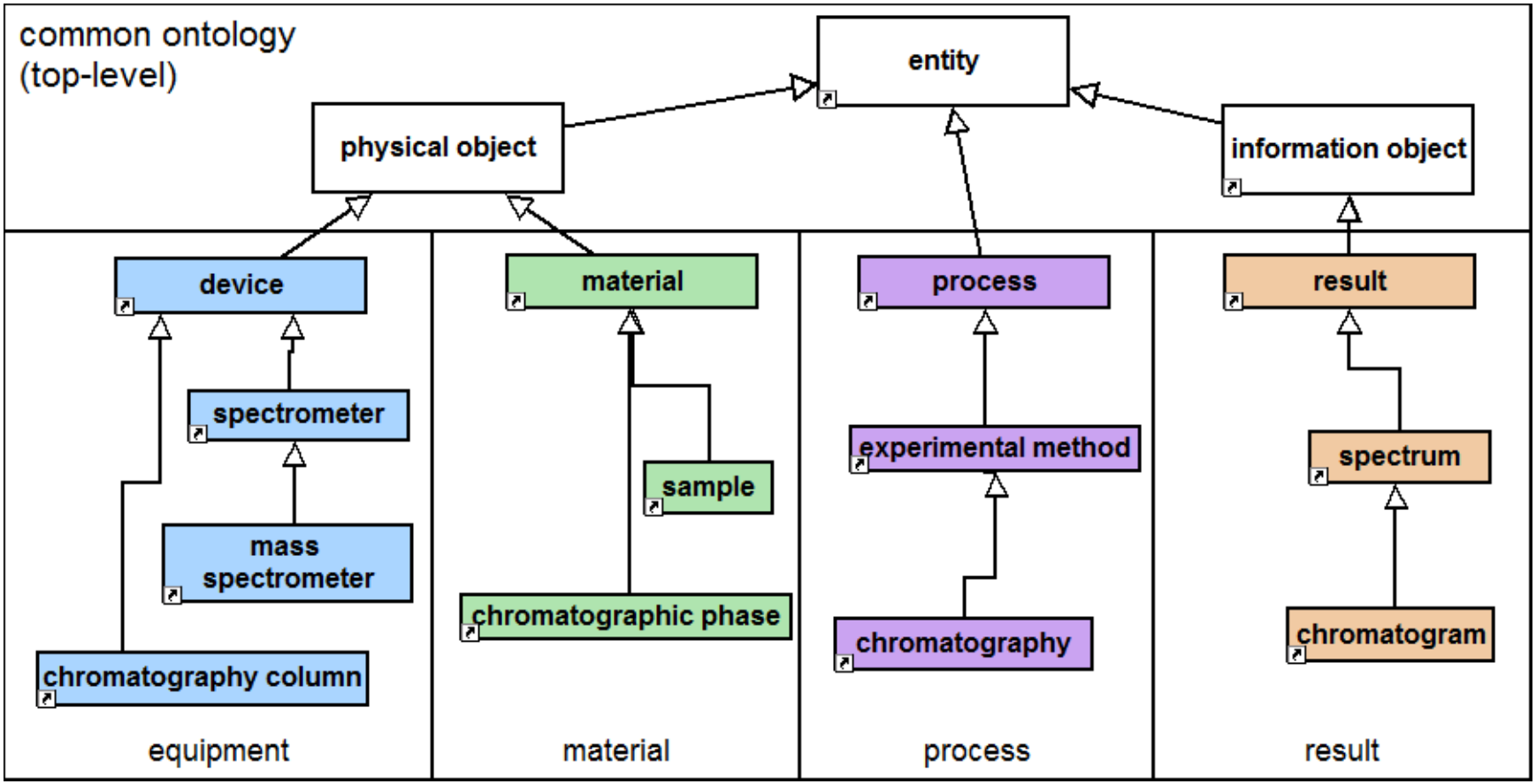
ADF is based on Hierarchical Data Format (HDF 5), which is specificially designed to store and organize large amounts of numerical data.

API Stack

- The Allotrope Framework provides APIs to read and write data contained in ADF
- Thus, developers do not have to concern themselves with RDF, SPARQL, semantics or complex graph patterns.



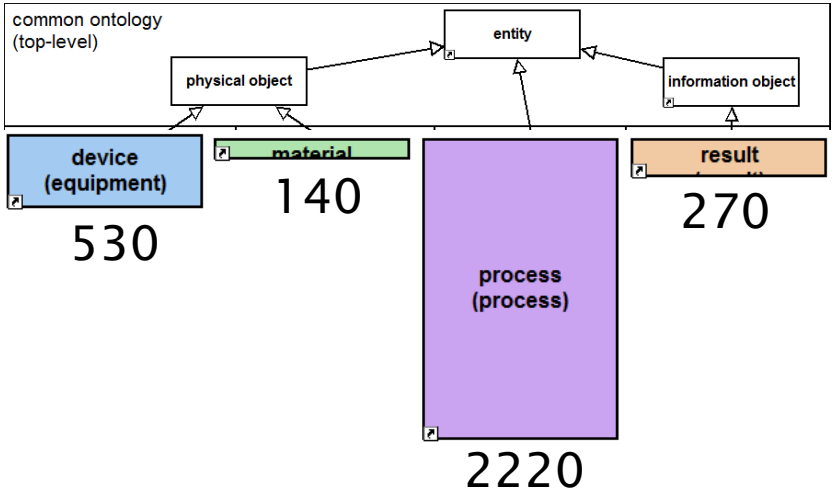
Allotrope Foundation Taxonomies (AFT)



Scope and Current Status

- 13 analytical techniques are already implemented:
  - small molecules:
    - gas chromatography
    - Karl Fischer
    - liquid chromatography
    - mass spectrometry
    - nuclear magnetic repulsion spectrometry
    - thermogravimetric analysis
    - ultra violet spectrometry
  - large molecules:
    - capillary electrophoresis
    - cell counter
    - cell culture analyzer
    - blood gas analysis
  - both:
    - balance
    - pH

Number of Classes:

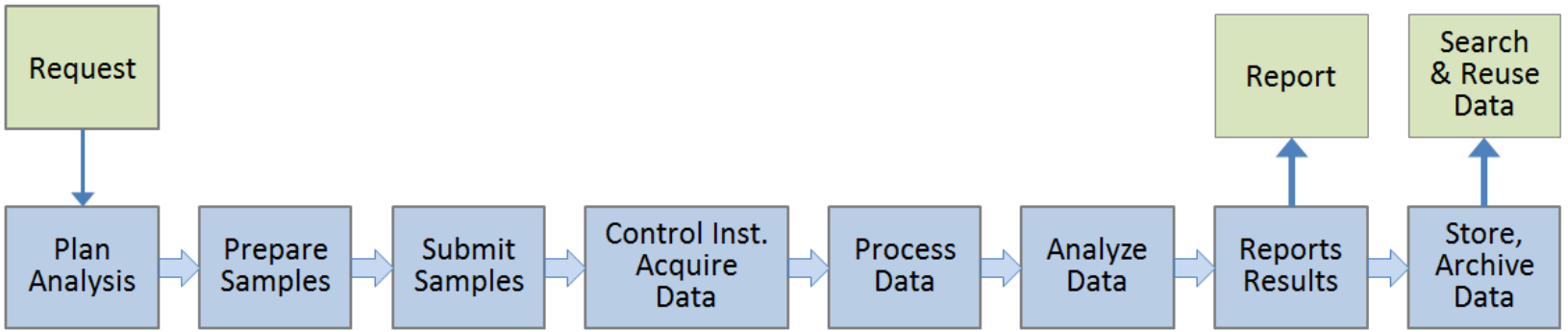


Reused Vocabularies and Ontologies

- Directly imported:
  - Simple Knowledge Organization System (SKOS)
  - Quantities, Units, Dimensions and Data Types Ontologies (QUDT)
  - The RDF Data Cube Vocabulary (QB)

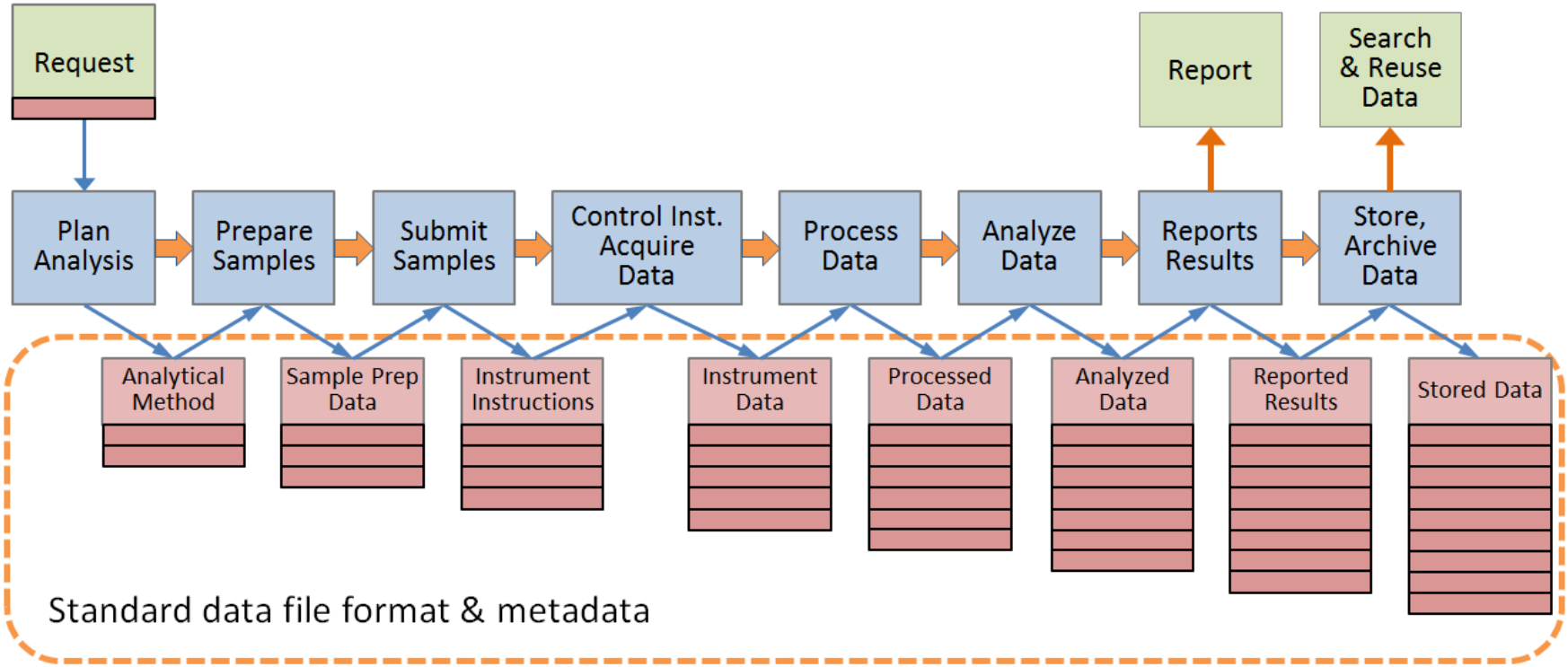
- Partly reused definitions:
  - Chemical Methods Ontology (CHMO)
  - Proteomics Standards Initiative – Mass Spectrometry (PSI-MS)
  - International Union of Pure and Applied Chemistry (IUPAC)
  - …

Analytical Workflow

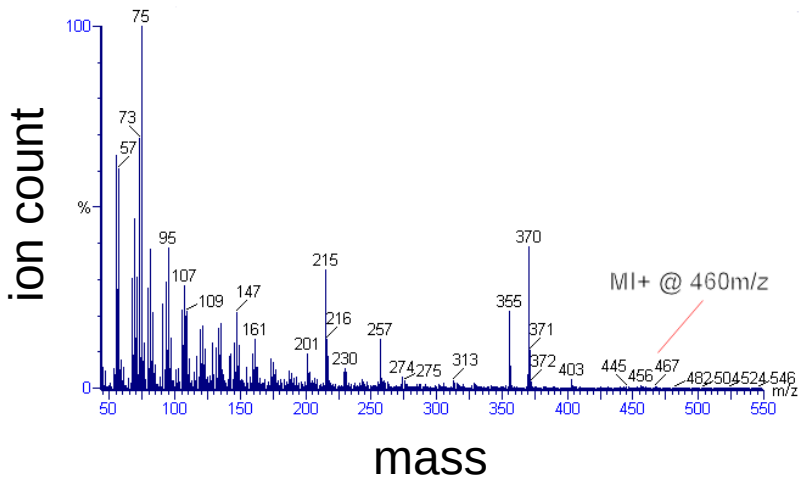


Analytical Workflow

- The basic analytical workflow and data flow gets standardized



Liquid Chromatography Mass Spectrometry



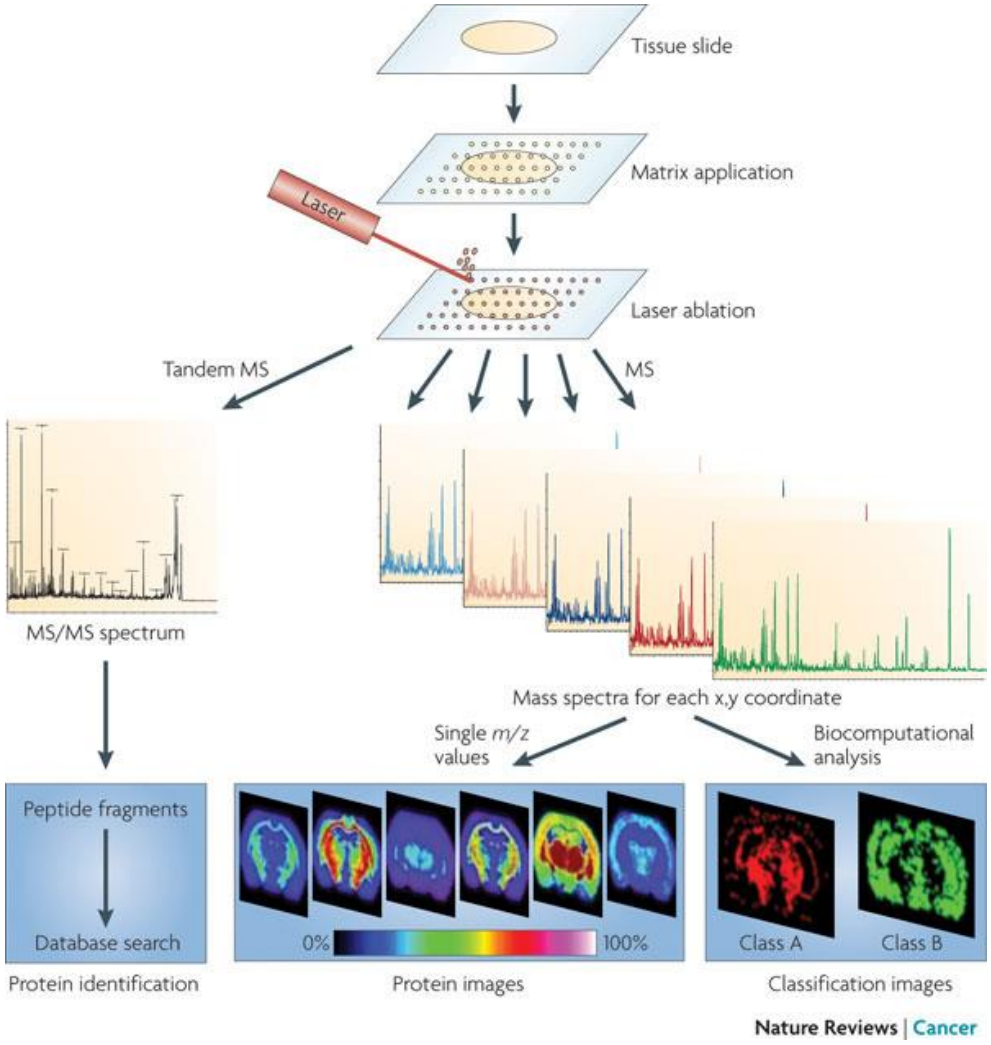
Data set of rank 2

Additional dimensions:
- sample
- retention time
- device
- …

Only meta data is expressed in RDF, while the numeric data is natively represented in HDF 5.

The **ADF Data Cube Ontology** provides the mapping between RDF meta data descriptions and physical storage in HDF 5.

Imaging Mass Spectrometry



Nature Reviews Cancer 10, 639-646 (September 2010) | doi:10.1038/nrc2917

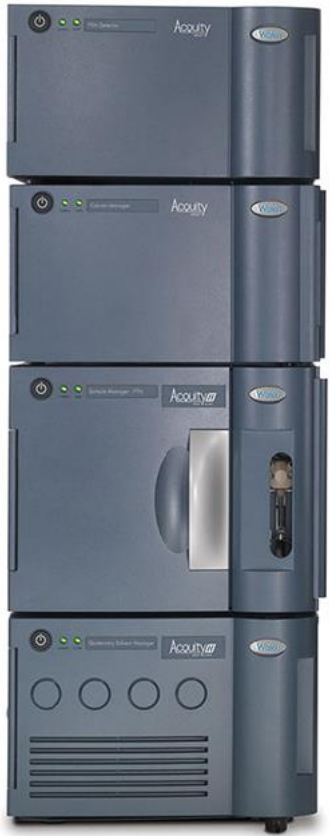
High Performance Liquid Chromatography

http://registry.mycompany.com/systems/hplc/hplc-uv/

Base URL in Registry

<HPLCSystem1>

af-e:has component



<HPLCSystem1/PDADetector>

<HPLCSystem1/ColumnManager>

<HPLCSystem1/SampleManager>

<HPLCSystem1/QuaternarySolventManager>

Linked Data Platform relative URLs under HPLC-UV

Conclusion

- **Initially:** Experiments were performed to get approval for drugs.
- **Today:** Experiments generate data that can be used in many different contexts.

Why Semantics?

- Good framework for *standardized* data descriptions and needed to realize the potential of the available data
- Linked Data allows to relate information stored in ADF with additional context: e.g. materials, devices, chemicals, processes, locations etc.

Questions?

Heiner Oberkampf
heiner.oberkampf@osthus.com
www.osthus.com

Allotrope Foundation:
www.allotrope.org