Allotrope Foundation

Building an Open Framework for Laboratory Data

Pharmaceutical companies collaborate on building a Framework to manage Analytical Data more efficiently

Dr. Gerhard Noelken

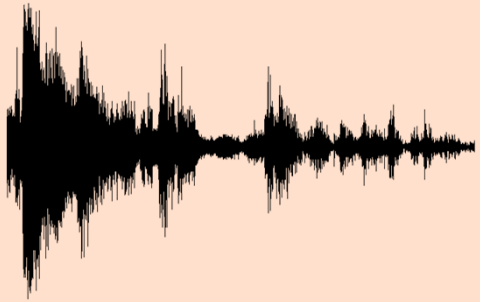
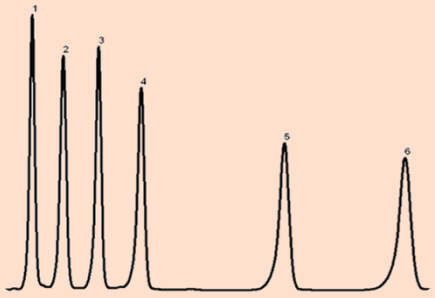
Allotrope BoD member, Pfizer Allotrope Liaison

GINAS Symposium, Uppsala

7 September 2015

- **MOTIVATION**
- **THEORY**
- **REDUCING IT TO PRACTICE**

What is the problem we are trying to solve?

	DATA SOURCE	DATA	VALUE FROM DATA
MUSIC			
DATA	Electronic Lab Notebook Chromatography Data System Data Archive		

Why is access to music so much easier than access to scientific data?

Think about music...

Music is typically stored in a small number of *standard, non-proprietary* formats…

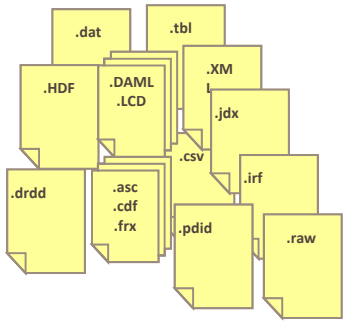
…with contextual *metadata* that are *complete, consistent & correct*

Artist	**Genre**
Album	**Date**
Song	**Artwork**

…enabling the user to *find, share* and *enjoy* it years later from any device *easily*!

Think about scientific data...

Scientific data is typically stored in a wide variety of *non-standard, proprietary* formats…



…with contextual *metadata* that are *hard to find and sometimes inconsistent*

Material	**Process**
Equipment	**Result**

…making it *costly and sometimes difficult* to find and get value from it.

What if scientific data was as easy to access as music?

If we...

- Store scientific and process data in a *standard format* with *contextual metadata* that is...
  - correct
  - complete
  - consistent
  - compliant

We could...

- *Find data in seconds.*
- *Be confident* that the *data* that underpins our *decisions* is *accurate*, *complete*, and *compliant*.
- *Build data quality and data integrity into the system*, eliminating the need for many SOPs and quality investigations.
- *Simplify, automate* and *improve* laboratory and manufacturing processes.
- *Automatically create technical reports*, audit trails, and substantial portions of regulatory submission documents.
- *Answer complex questions*, not just those accessible via simple queries - by linking data from diverse, disparate sources.

Allotrope Foundation

Member Companies

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

Secretariat

Drinker Biddle

- Project Management
- Legal & Logistical Support

Professional Software Firm

OSTHUS
success with R&D

- Framework Development
- Technical Leadership

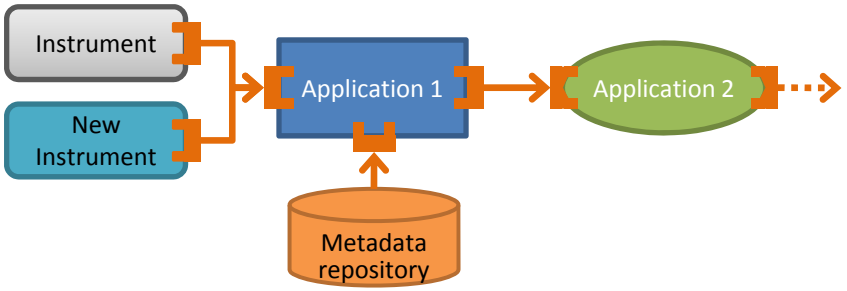
Partner Network

Allotrope Partner Network

ACD/Labs	Mettler Toledo	Waters
Agilent	Persistent	
Biovia	Riffyn	Erasmus Univ. Med Center
BSSN	Sartorius	University of Southampton
IDBS	Shimadzu	
Mestrelab Research	Thermo Scientific	

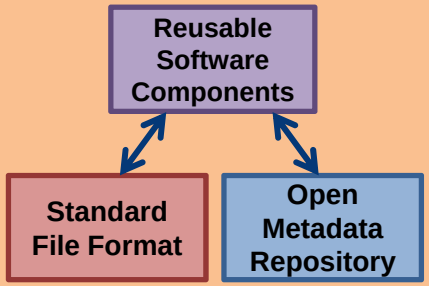
- MOTIVATION
- **THEORY**
- REDUCING IT TO PRACTICE

What is Allotrope Creating?

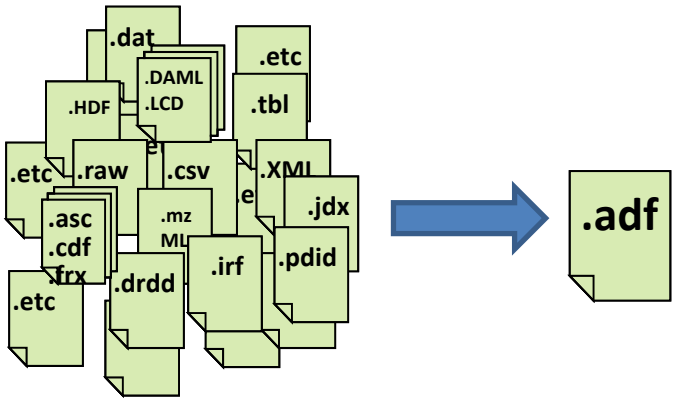


A **toolkit** that enables use of the standards & metadata in software development

Allotrope Foundation Framework



File format for any technique or instrument



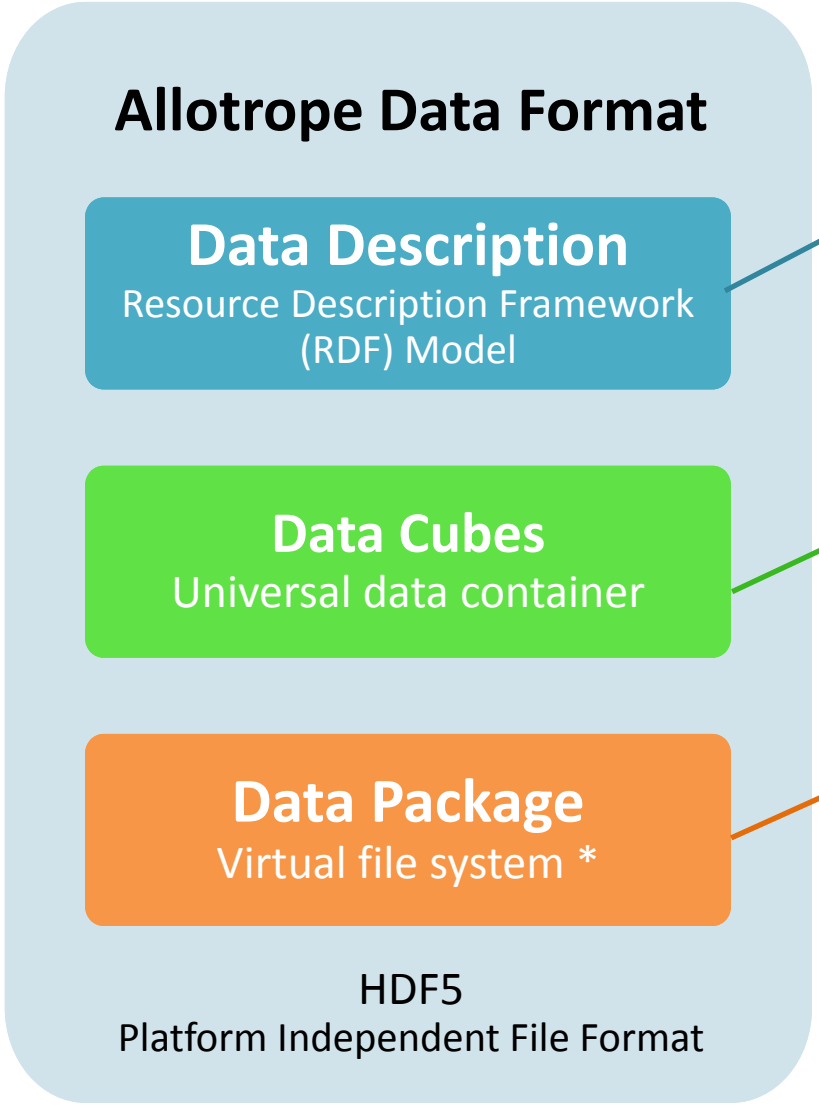
Standard vocabulary & structure for metadata

Project	Test	Instrument
AF 0012354	IR Fingerprinting	QC Lab #33B 380 FT-IR
AE0012764	Bulk & Tapped Density	ASTM Standard Seive #6
AF 12989	NMR Characterization	AM500
	Tapped & Bulk Density	Sieve XXX
AF0045674	Caractérisation RMN	Nouvelle DRX600
AF-0034558	IR	iS10 FT-IR

With the Metadata Repository

Project	Test	Instrument
AF0012354	IR Fingerprinting	380 FTIR/-SN/145453
AF0012764	Bulk and Tapped Density	ASTM Sieve-SN/3452
AF0012989	NMR Characterization	AM500-SN/0034578
AF0013142	Bulk and Tapped Density	ASTM Sieve-SN/09783
AF0045674	NMR Characterization	DRX600-SN/10234567
AF0034558	IR Fingerprinting	iS10 FTIR/-SN/341980

ADF: a universal file format for scientific data



Contains semantic descriptions of:

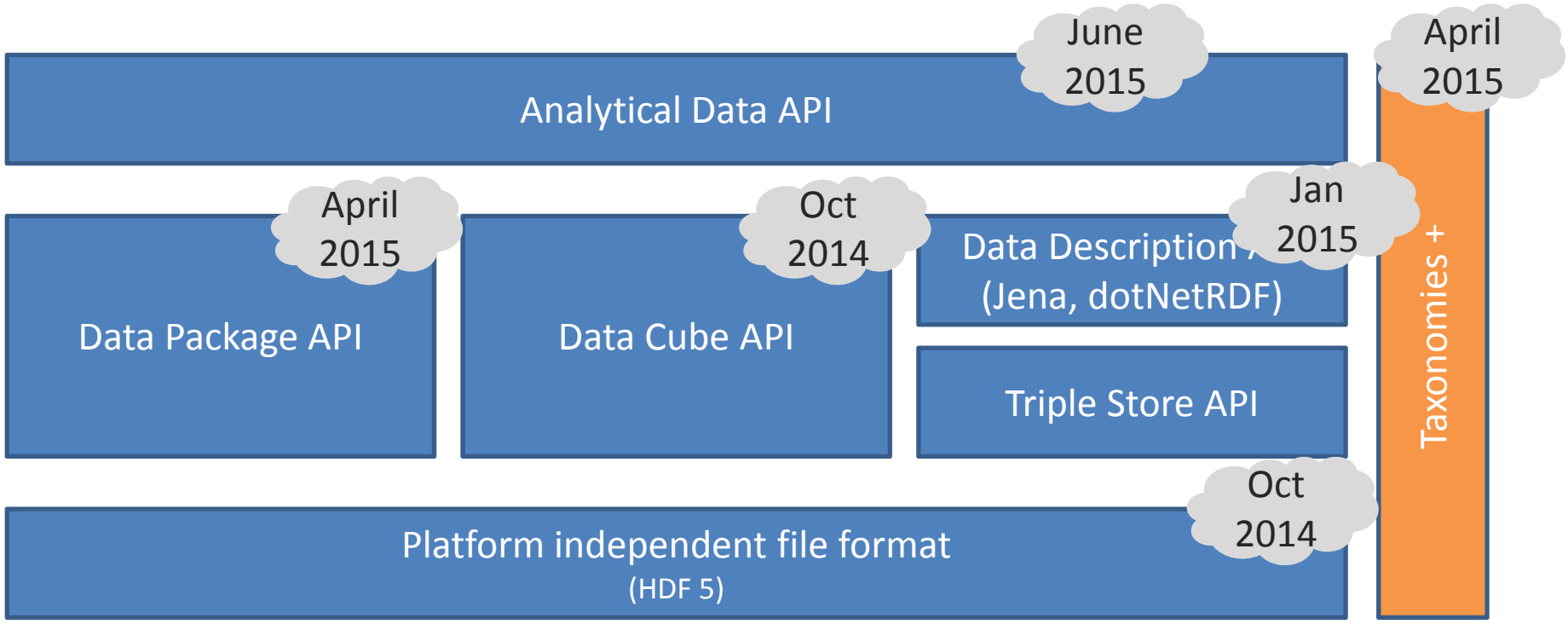
- Method, instrument, sample, process, result, etc.
- Data cube metadata
- Binary file metadata

Analytical data represented by one- or multidimensional arrays

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

** Use is optional*

ADF Class Libraries + Decoupled Taxonomies



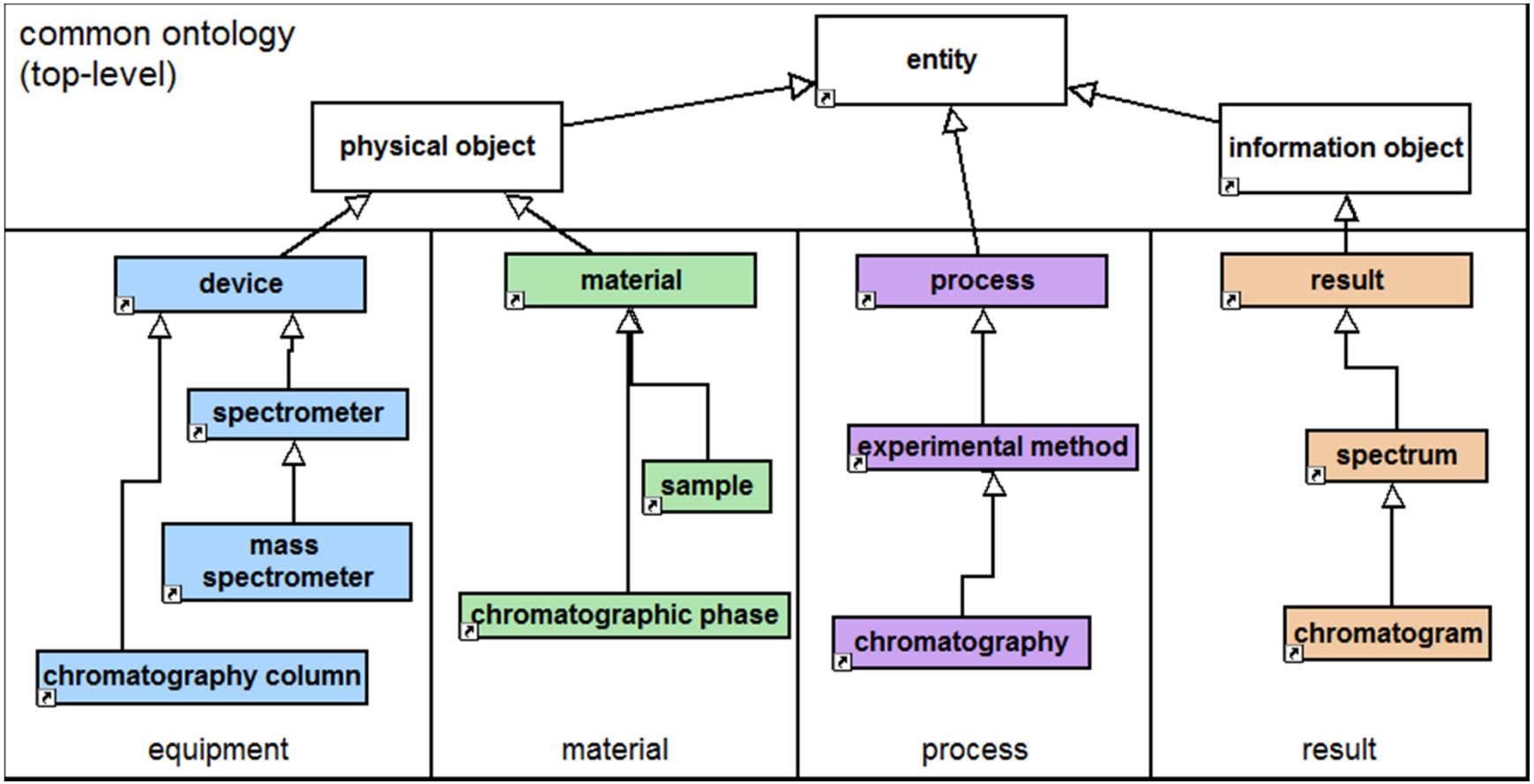
Languages: Java, C#

Allotrope Taxonomies: An Extensible Metadata Model

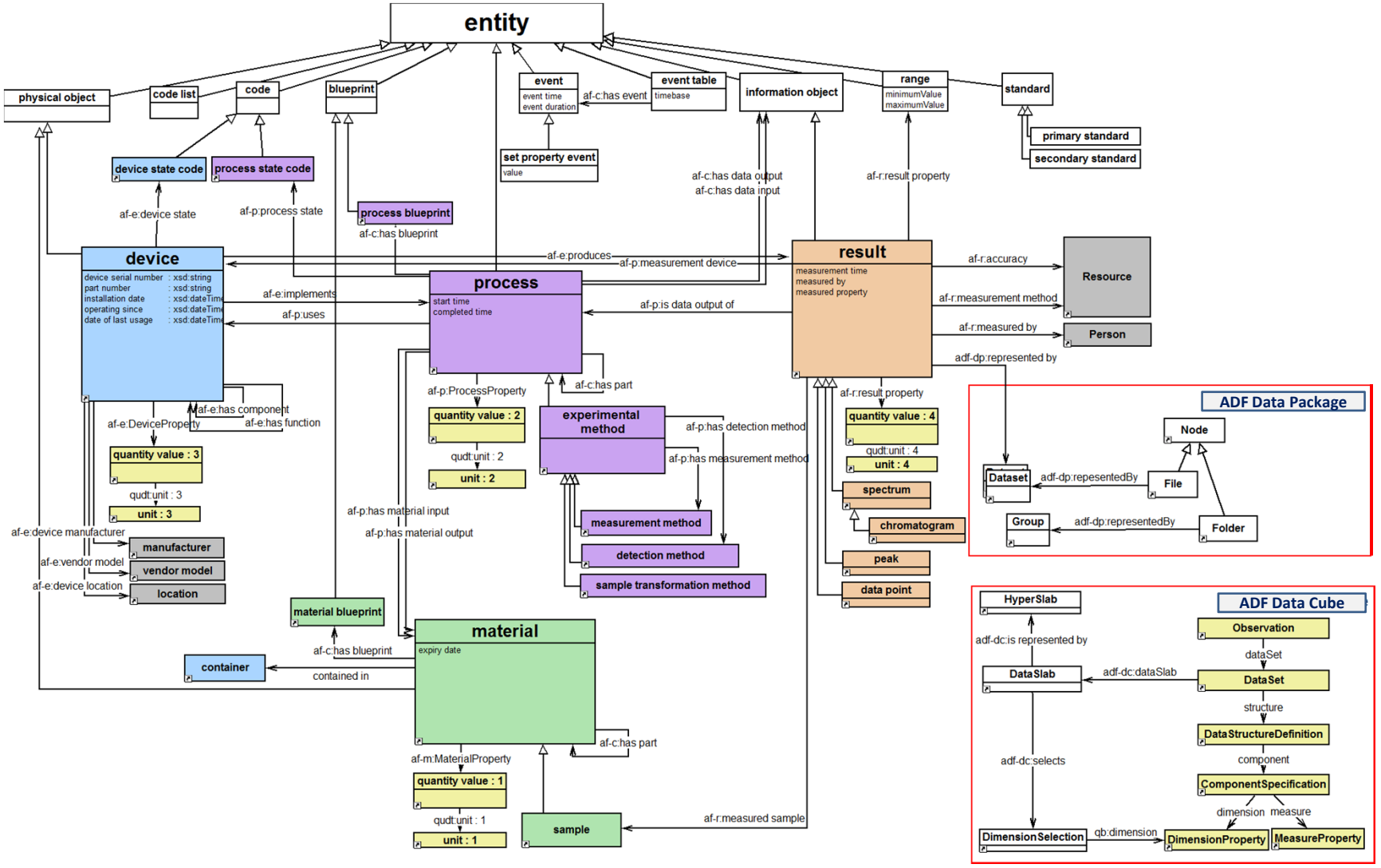
- **A library of extensible taxonomies**
  - Uses W3C standards
  - Easy to understand and maintain by SMEs and Vendors
- **Start by harvesting existing available concepts**
  - PSI-MS; IUPAC; RSC Chemical Methods Ontology; Dictionary of weighing terms; AnIML, etc
- **Reproducible & efficient collaboration model**
  - Leverages knowledge engineers & member company scientists
  - 2-3 weeks to develop initial version of a new taxonomy

- **Initial versions of 12 analytical techniques already implemented:**
  - gas chromatography
  - Karl Fischer
  - liquid chromatography
  - mass spectrometry
  - nuclear magnetic resonance spectroscopy
  - thermogravimetric analysis
  - ultra violet spectrometry
  - cell counter
  - cell culture analyzer
  - blood gas analysis
  - balance
  - pH

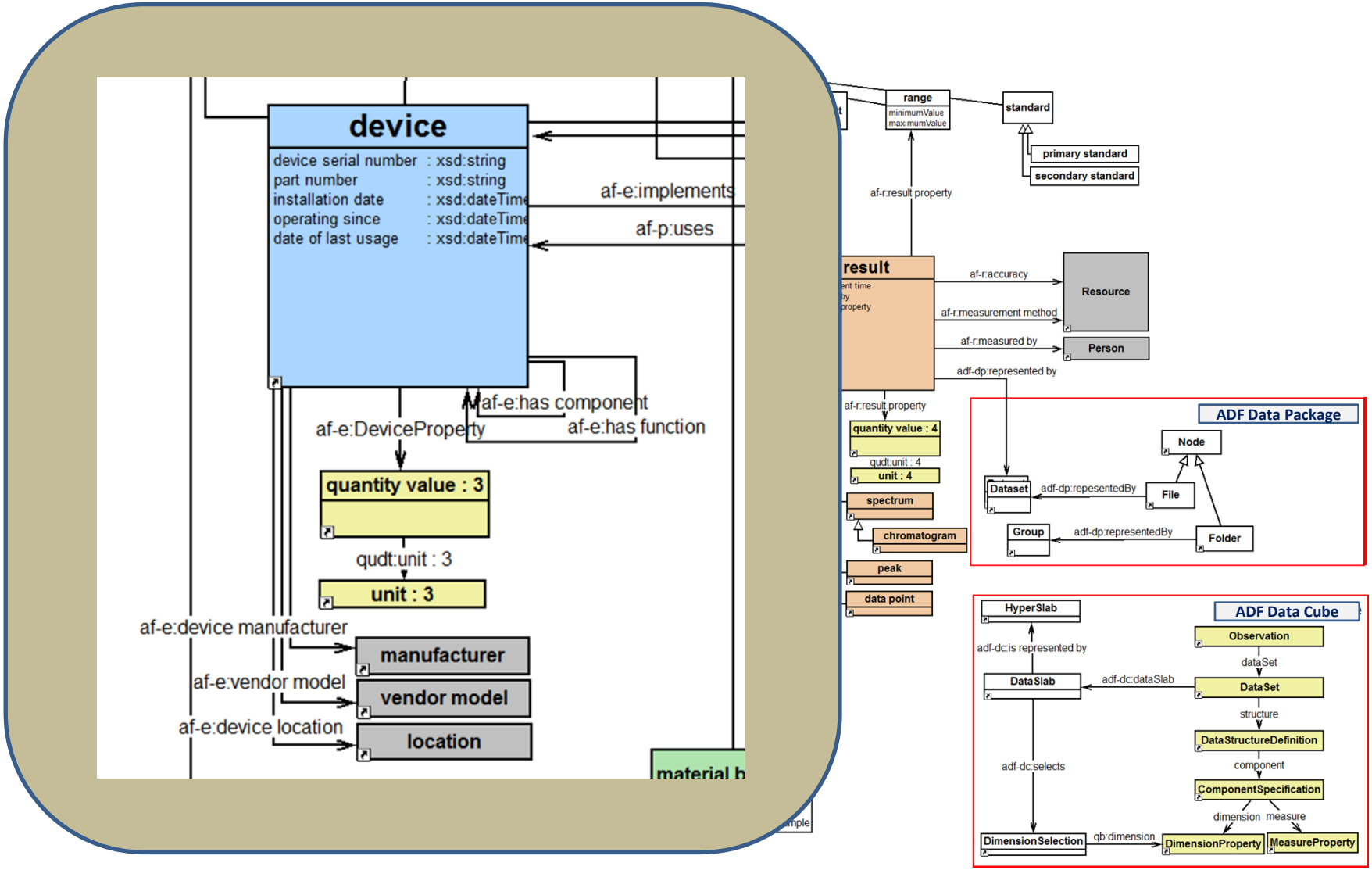
Allotrope Foundation Taxonomies



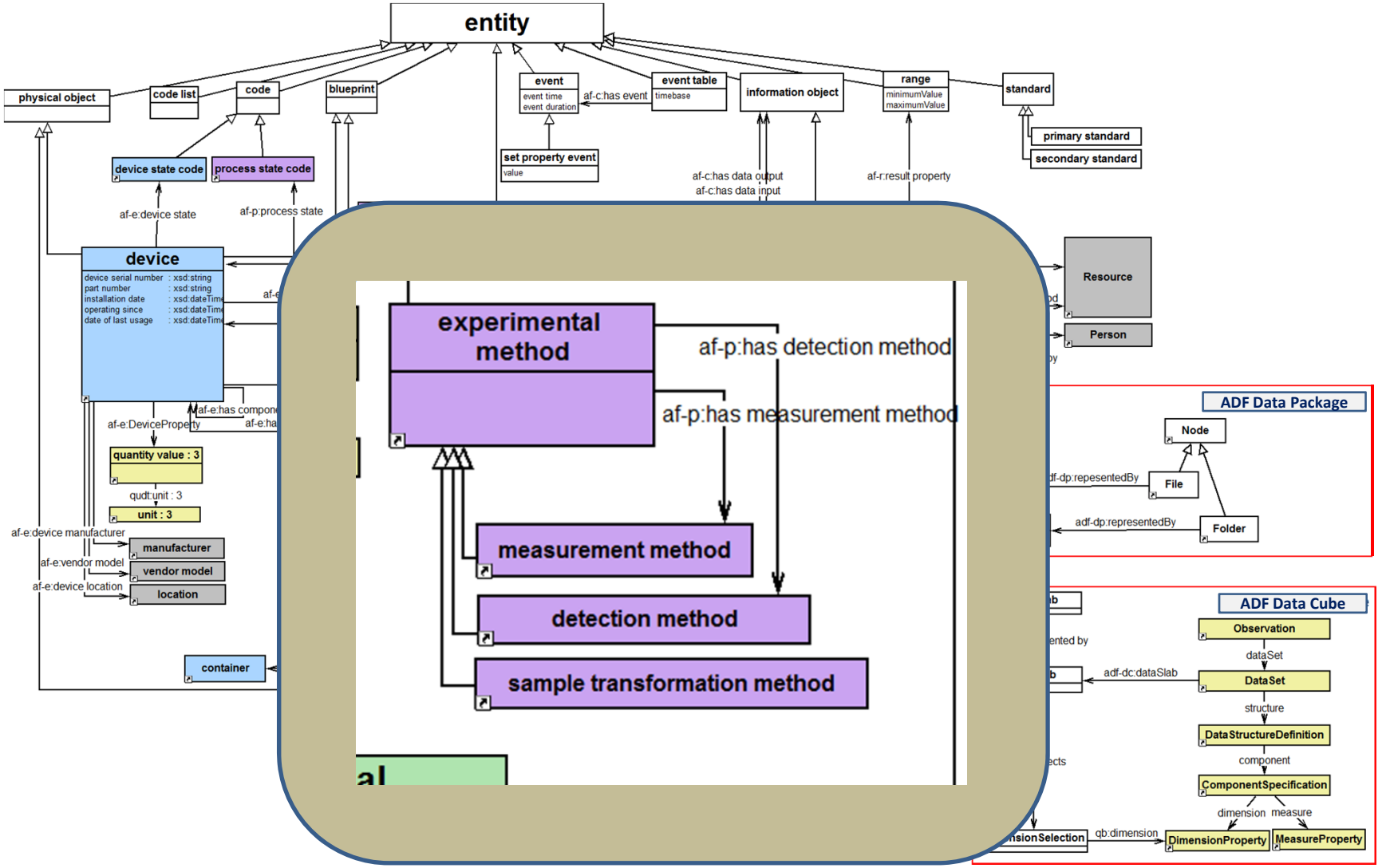
The Big Picture

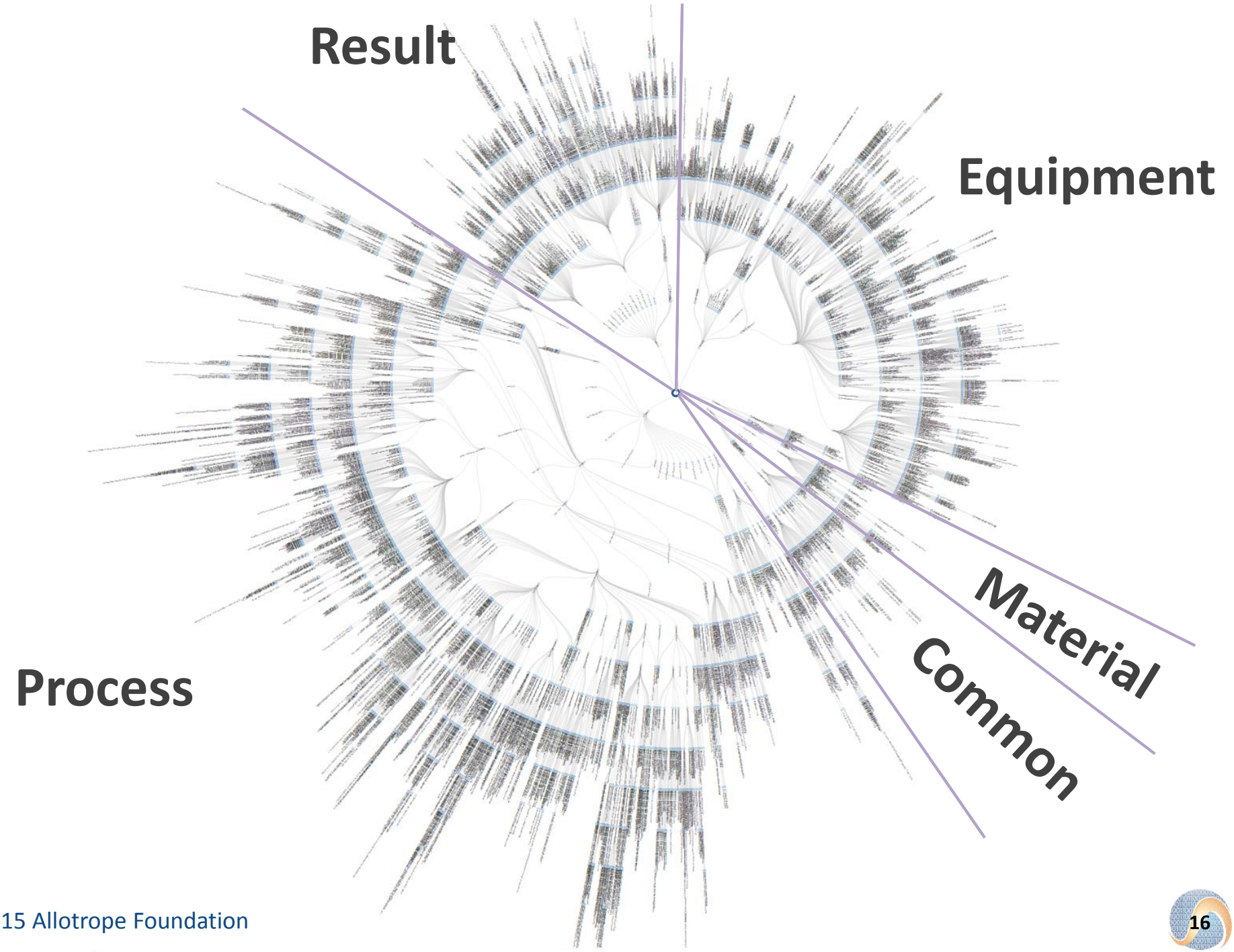


The Big Picture

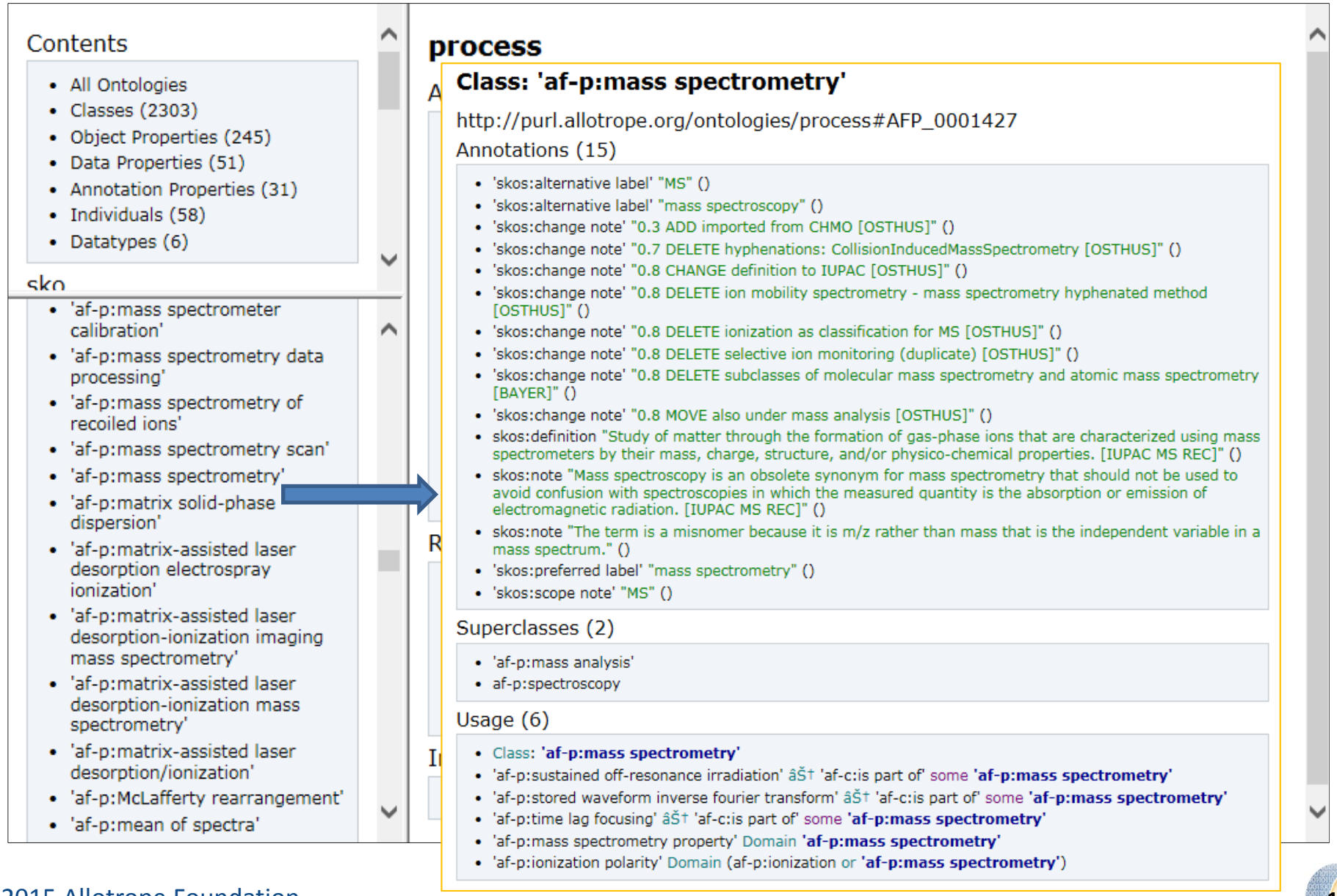


The Big Picture

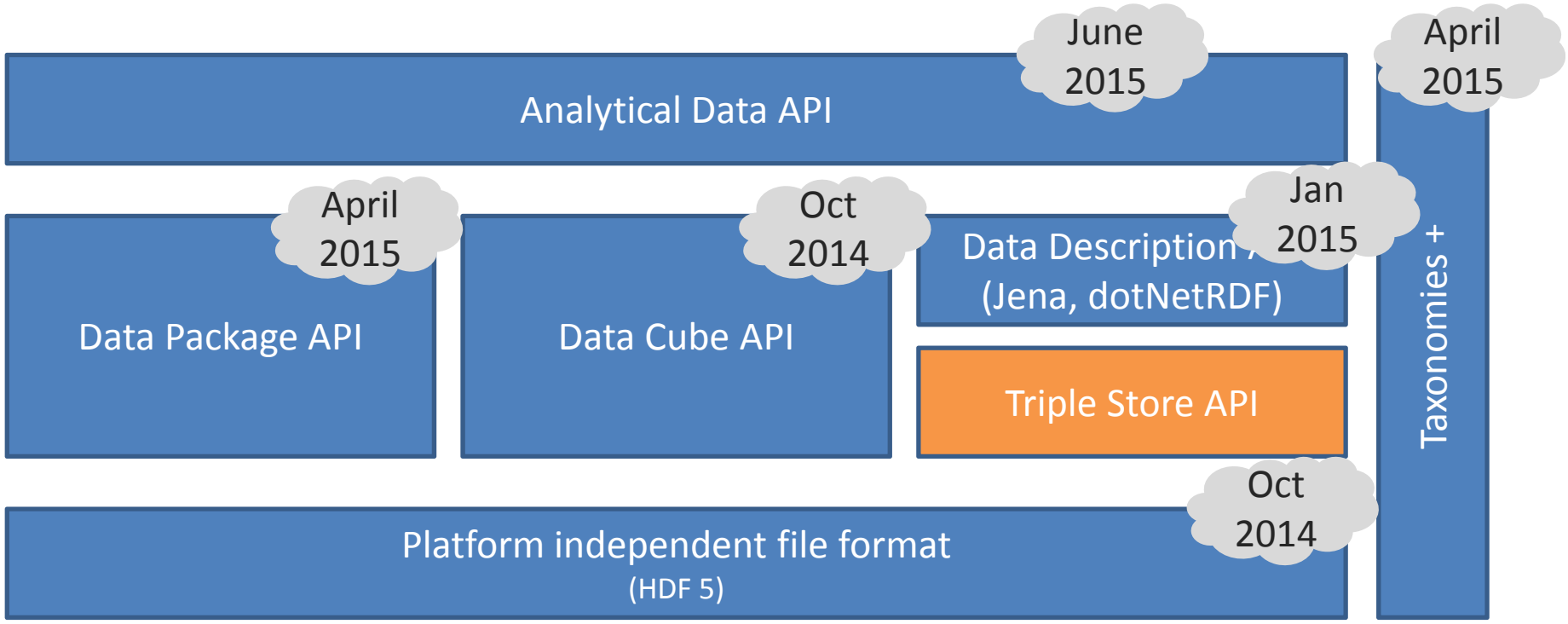




AF Taxonomies Documentation



ADF Class Libraries + Decoupled Taxonomies



Languages: Java, C#

RDF Data Model

- Subject-Predicate-Object (Triple)

- Example:

<Sample 1> type <Sample>
<Sample 1> createdOn ‘2015-03-13’
<Sample 1> createdBy <person X>
<Sample 1> hasBarcode ‘1234567890’

Data Shapes Constrain How We Use Taxonomies in the Real World

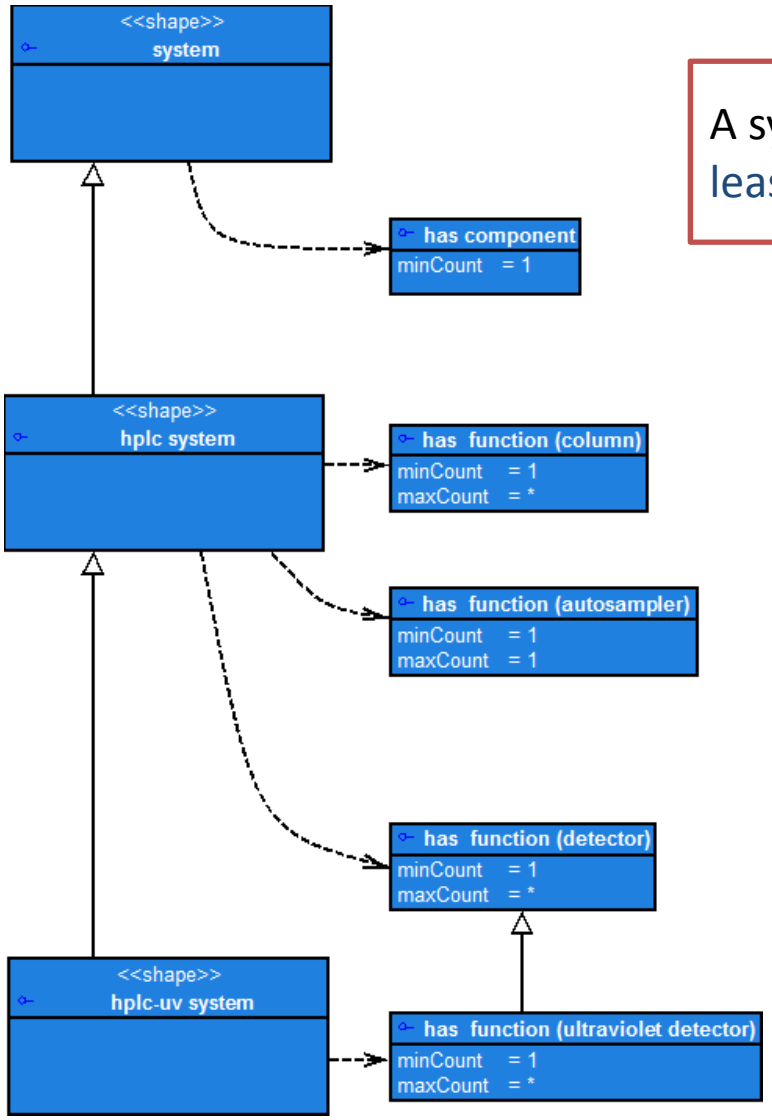
- Taxonomies provide an unconstrained vocabulary that we can use to describe things (instances) in our open world and give them a meaning (= **what** it is)

- We need a mechanism to define data structures (schemas, templates) that describe **how** to use the taxonomies for a given purpose in a standardized (= reproducible, predictable, verifiable) way

- Shapes Constraint Language (SHACL, expressed as RDF triples) is an emerging standard to do this

http://www.w3.org/2014/data-shapes/charter

Using Data Shapes: Equipment



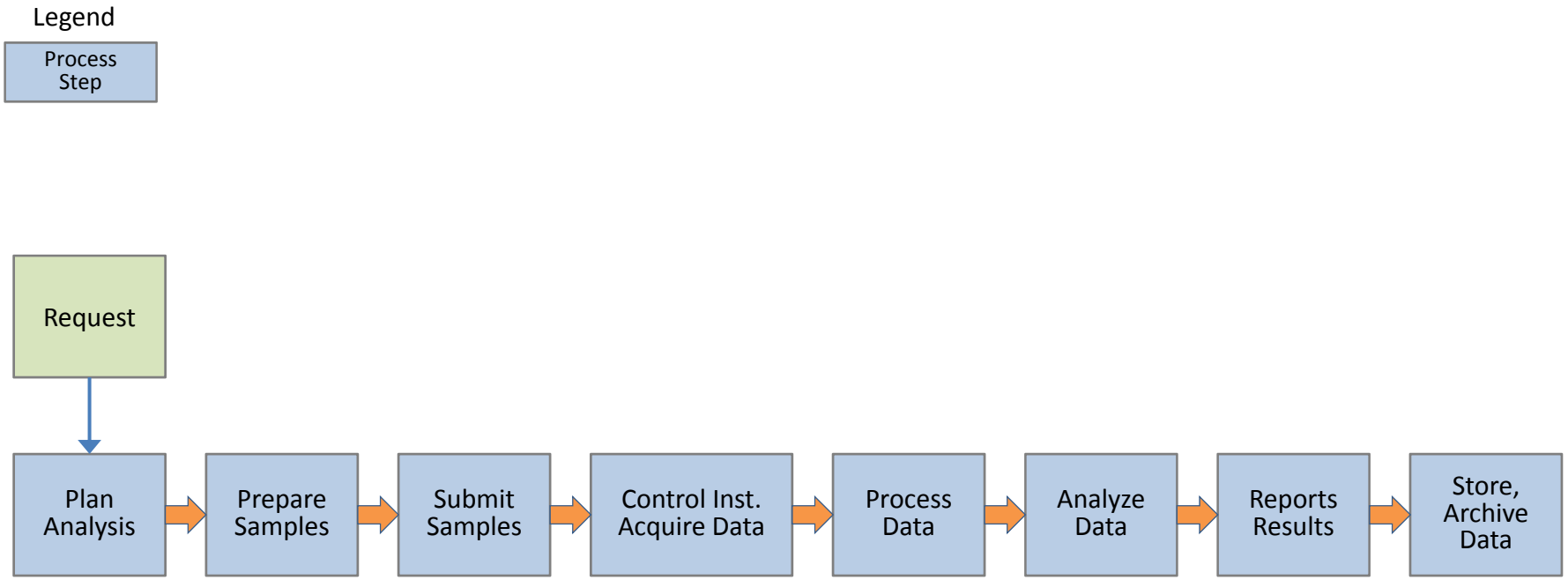
A system has at least 1 component

Shape hierarchies define **additional** constraints

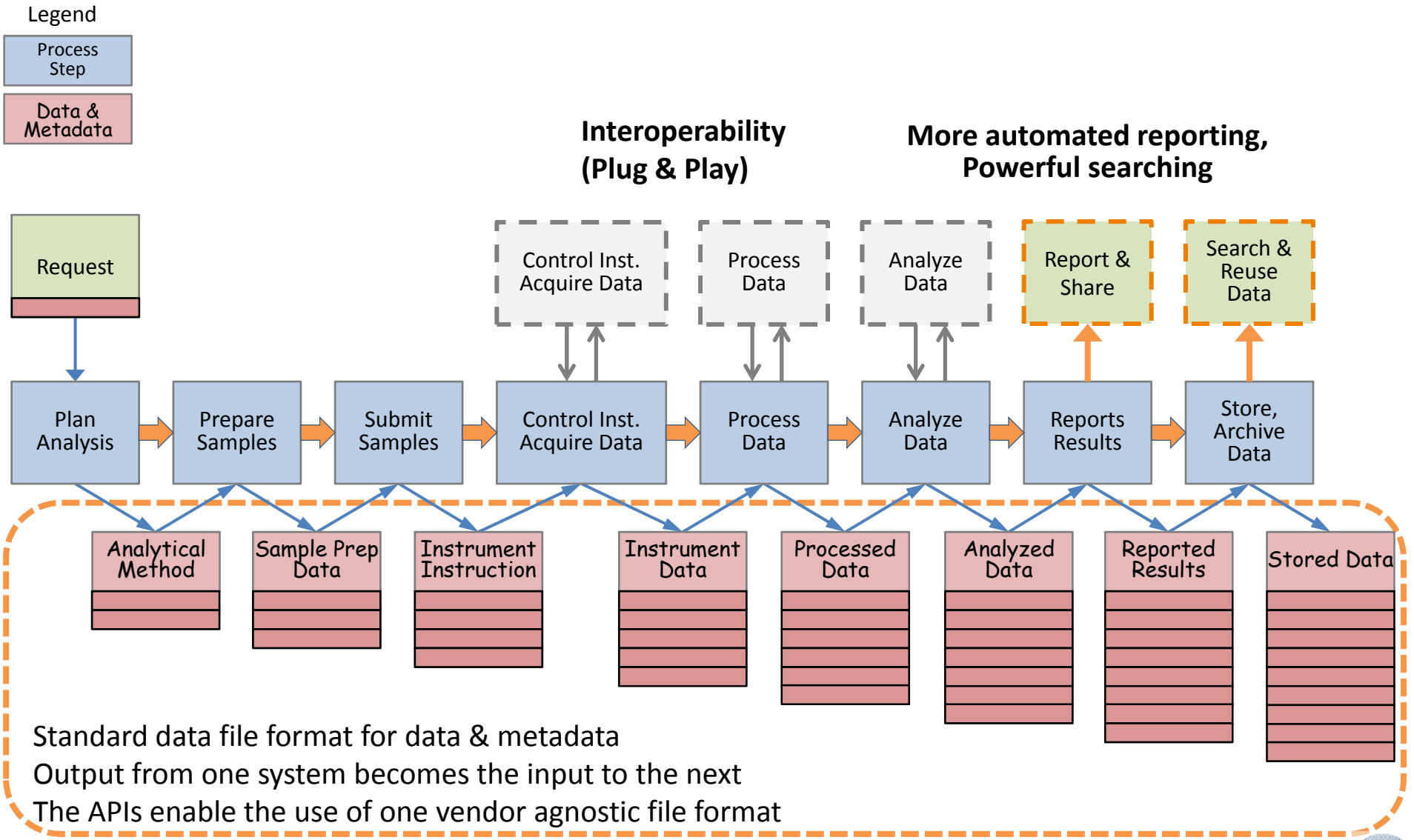
A hplc system has at least 1 component **and** has at least 1 column, exactly 1 autosampler and at least 1 detector

A hplc-uv system has at least 1 component **and** has at least 1 column, exactly 1 autosampler and at least 1 detector **and** at least 1 uv-detector

The ADF enables a self-contained documentation of the data & metadata



The ADF enables a self-contained documentation of the data & metadata



Allotrope Data Format

Data Description

Request

```
<myAnalysisRequest> a adf-p:AnalysisRequest ;
    adf-p:Input <S0815>;
    adf-p:hasPurpose <drugSubstanceReleaseTesting> .

<S0815>a adf-m:Sample ;
  adf-m:ChemicalSubstance <Aspirin> ;
  adf-c:Barcode "S0815" .
```

Sample prep

```
<S0815-1-2> a adf-m:Sample, adf-m:Solution ;
            adf-m:createdFrom <S0815-1-1>;
            adf-m:SampleVolume  "50.0  ml";
            adf-m:SampleDiluent <AcCN> ;
            adf-m:Solute <S0815-1-1>;
            adf-m:Stability adf-m:Ambient,
                            adf-m:ProtectFromLight ;
            adf-m:MaxStorageDuration "P7D"^^xsd:duration ;
            adf-m:SampleConcentration "0.2 g/l";
            adf-m:SampleContainer  [
                    a adf-e:VolumetricFlask ;
                adf-e:ContainerLabel "S0815 in Acetonitrile" ;
                ghs:Pictogram ghs:F, ghs:Xn ;
                ghs:HazardStatement ghs:H225 , ghs:H302, ... ;
                    adf-c:Barcode "1001231231";
                adf-e:ContainerVolume "50.0 ml" ;
            ];
.

<AcCN> a adf-m:Substance .

<S0815-1-1> adf-m:consumed true .
```

Method

```
<Method007/HPLC> a adf-p:HighPerformanceLiquidChromatography ;
                 adf-p:InputSpecification [ a adf-m:Sample, ldom:Shape ] ;
                 adf-p:ResultSpecification [ a adf-r:Chromatogram ,
ldom:Shape ] ,
                                           [ a adf-r:UltravioletSpectrum, ldom:Shape ] ;
                 adf-p:hasEventTable <Method007/HPLC/gradientTable > ;
                 adf-p:hasMobilePhase  <mp1>, <mp2>, <mp3> ;
                 adf-p:hasStationaryPhase  <columnPhase> ;
                 adf-p:hasDetection <Method007/HPLC/UV>,
                                    <Method007/HPLC/PDA> ;
.
<Method007/HPLC/gradientTable> a adf-p:EventTable .

<Method007/HPLC/UV> a adf-p:UltravioletDetection ;
            adf-p:DetectionWavelength "220 nm" ;
            adf-p:DetectionResolution  "2.0 nm" .

<Method007/HPLC/PDA> a adf-p:PhotodiodeArrayDetection ;
            adf-p:DetectionWavelength [
                a adf-c:Range ;
                adf-c:Minimum  "190 nm" ;
                    adf-c:Maximum  "400 nm " ;
            ];
```

Instrument instruct.

```
<HPLCR/eventTable1> a adf-p:EventTable;
                    adf-p:hasEvent
[ adf-p:EventTime 0:  adf-p:Subject <columnValve>;
adf-p:Property adf-p:ColumnSelection; adf-p:Value <column1>  ],
[ adf-p:EventTime 0: adf-p:Subject <lamp>:  adf-p:Property adf-p:Status; adf-p:Value adf-p:StatusOn ],

[ adf-p:EventTime 0: adf-p:Subject <solventValve>; adf EventProperty adf-p:Status; adf-p:Value <B5>];

[ adf-p:EventTime 0: adf-p:Subject <A>: adf-p:Property adf-p:Composition; adf-p:Value "80%" ],
[ adf-p:EventTime 0: adf-p:Subject <B>: adf-p:Property adf-p:Composition; adf-p:Value "20%"],
[ adf-p:EventTime 0: adf-p:Subject <eluent>: adf-p:Property adf-p:FlowRate adf-p:Value "0.3ml/min"],

[ adf-p:EventTime 1: adf-p:Subject <A>: adf-p:Property adf-p:Composition; adf-p:Value "80%" ],
[ adf-p:EventTime 1: adf-p:Subject <B>: adf-p:Property adf-p:Composition; adf-p:Value "20%"],
[ adf-p:EventTime 1: adf-p:Subject < eluent >:  adf-p:Property adf-p:FlowRate; adf-p:Value  "0.2ml/min"],

[ adf-p:EventTime 3: adf-p:Subject <A>: adf-p:Property adf-p:Composition; adf-p:Value "0%" ],
[ adf-p:EventTime 3: adf-p:Subject <B>:  adf-p:Property adf-p:Composition; adf-p:Value "100%"],

[ adf-p:EventTime 5: adf-p:Subject <A>: adf-p:Property adf-p:Composition; adf-p:Value "80%" ],
[ adf-p:EventTime 5: adf-p:Subject <B>: adf-p:Property adf-p:Composition; adf-p:Value "20%"],
[ adf-p:EventTime 5: adf-p:Subject <lamp>:  adf-p:Property  adf-p:Status; adf-p:Value adf-p:StatusOff ],

[ adf-p:EventTime 5.1: adf-p:Subject <B>: adf-p:Property adf-p:Composition; adf-p:Value "100%"],
[ adf-p:EventTime 5.1: adf-p:Subject <solventValve>: adf-p:Property  adf-p:Status; adf-p:Value <B6>];

[ adf-p:EventTime 6: adf-p:Subject <B>: adf-p:Property adf-p:Composition; adf-p:Value "100%" ],
```

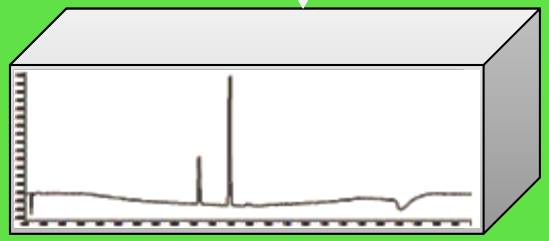
Data acquisition

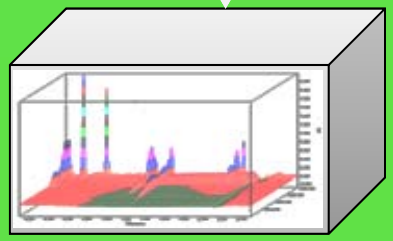
```
<myHPLC/1> a Run ;
...
adf-p:Result   <myHPLC/1/chromatogram>, <myHPLC/2/chromatogram> .

<myHPLC/1/chromatogram> a qb:Dataset, adf-r:Chromatogram ;
        adf-r:measurementDevice    <HPLC4/UVDetector> ;
        adf-r:measuredBy <JaneD> ;
        adf-r:measurementTime 2015-08-24T14:06:55+1;
        adf-qb:representedBy <hdf5:datacubes/2D HDF> ;
        qb:structure <Chromatogram2DCube> .

<myHPLC/2/chromatogram> a qb:Dataset, adf-r:Chromatogram ;
        adf-r:measurementDevice    <HPLC4/PDA> ;
        adf-r:measuredBy <JaneD> ;
        adf-r:measurementTime 2015-08-24T14:06:55+1;
        adf-qb:representedBy <hdf5:datacubes/3D HDF> ;
        qb:structure <Chromatogram3DCube> .

<Chromatogram2DCube>   a qb:DataStructureDefinition;
  qb:component [  qb:dimension adf-r:Time ],
                  [  qb:measure   adf-r:Absorbance; ].

<Chromatogram3DCube>   a qb:DataStructureDefinition;
  qb:component [  qb:dimension adf-r:Time; qb:order 1 ],
                  [  qb:dimension adf-r:Wavelength; qb:order 2],
                  [  qb:measure   adf-r:Absorbance; ].
```

Data Cubes

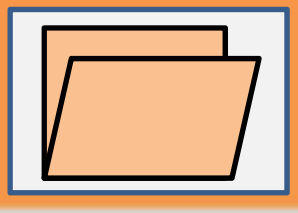
Chromatogram: 2D HDF

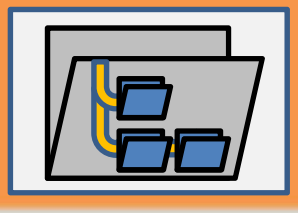
Chromatogram: 3D HDF





Data Package(s) (optional)

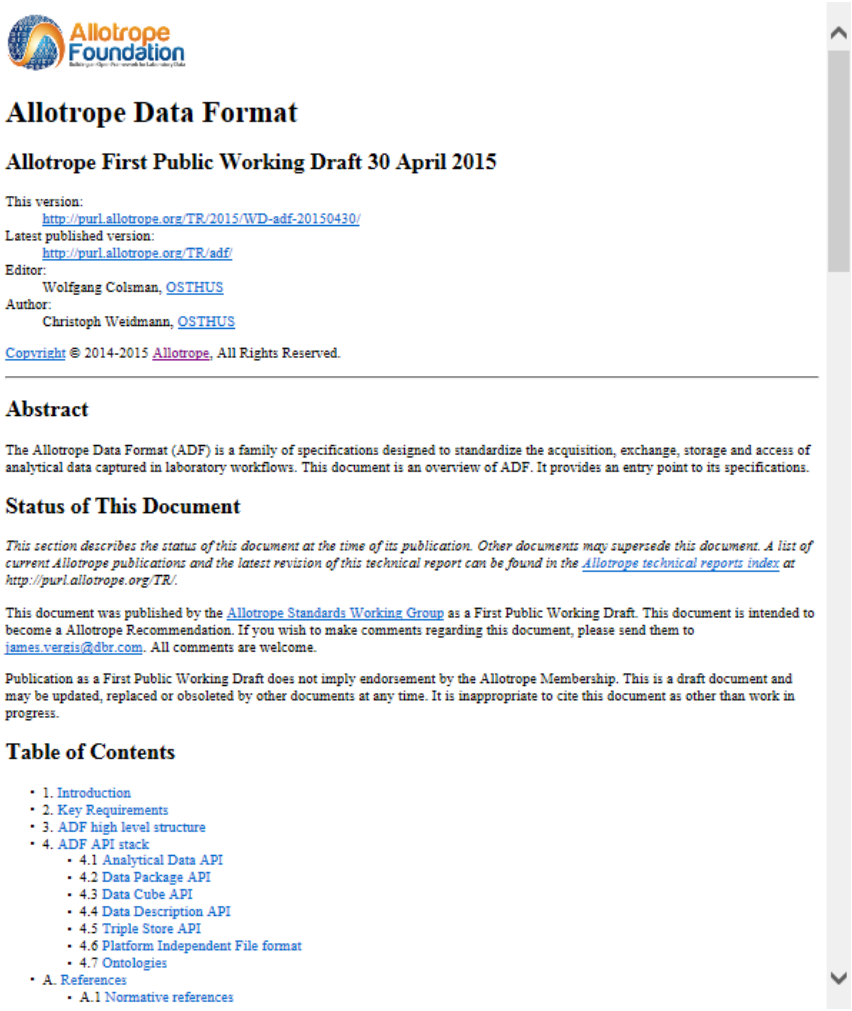




- MOTIVATION
- THEORY
- **REDUCING IT TO PRACTICE**

Alpha & Beta Release available to AF & APN members now

- **Taxonomies (work in progress)**
  - Mature versions for:
    - MS, LC, pH, weighing, UV
  - Initial versions of:
    - NMR, cell counter, blood gas, capillary electrophoresis, cell culture analyzer, thermogravimetric analysis, Karl Fischer, GC,
  - OWL, OWL + SKOS, Excel and OWLDoc formats
- **APIs**
  - Data Package, Data Description, Data Cube, Analytical Data
- **ADF & API documentation**
- **New Example Applications** for the ADF APIs



Allotrope Foundation

Allotrope Data Format

Allotrope First Public Working Draft 30 April 2015

This version:
http://purl.allotrope.org/TR/2015/WD-adf-20150430/
Latest published version:
http://purl.allotrope.org/TR/adf/
Editor:
Wolfgang Colsman, OSTHUS
Author:
Christoph Weidmann, OSTHUS

Copyright © 2014-2015 Allotrope, All Rights Reserved.

Abstract

The Allotrope Data Format (ADF) is a family of specifications designed to standardize the acquisition, exchange, storage and access of analytical data captured in laboratory workflows. This document is an overview of ADF. It provides an entry point to its specifications.

Status of This Document

This section describes the status of this document at the time of its publication. Other documents may supersede this document. A list of current Allotrope publications and the latest revision of this technical report can be found in the Allotrope technical reports index at http://purl.allotrope.org/TR/.

This document was published by the Allotrope Standards Working Group as a First Public Working Draft. This document is intended to become a Allotrope Recommendation. If you wish to make comments regarding this document, please send them to james.vergis@dbr.com. All comments are welcome.

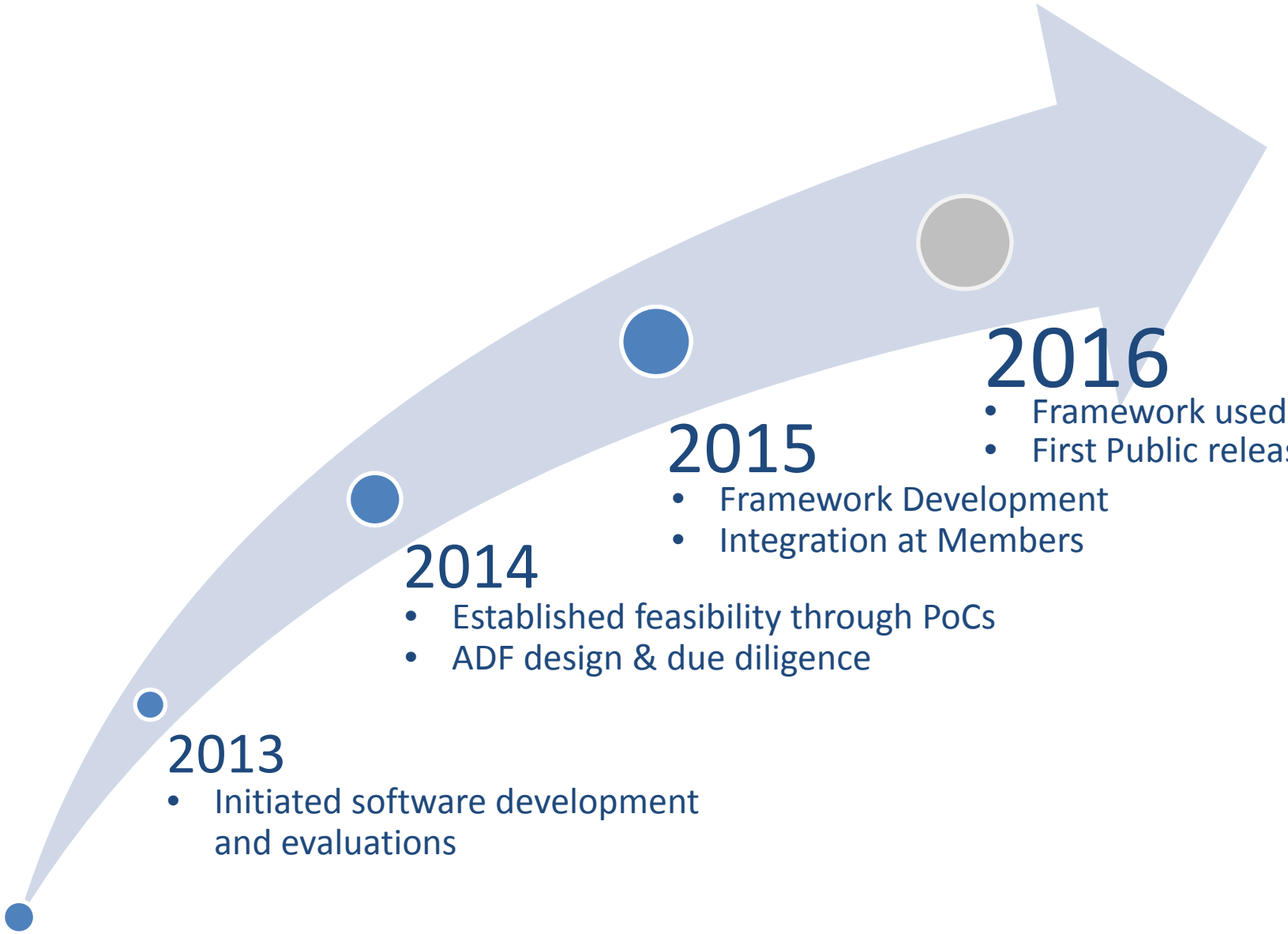
Publication as a First Public Working Draft does not imply endorsement by the Allotrope Membership. This is a draft document and may be updated, replaced or obsoleted by other documents at any time. It is inappropriate to cite this document as other than work in progress.

Table of Contents

- 1. Introduction
- 2. Key Requirements
- 3. ADF high level structure
- 4. ADF API stack
  - 4.1 Analytical Data API
  - 4.2 Data Package API
  - 4.3 Data Cube API
  - 4.4 Data Description API
  - 4.5 Triple Store API
  - 4.6 Platform Independent File format
  - 4.7 Ontologies
- A. References
  - A.1 Normative references

Current Allotrope Integration Project Dashboard

Project Type	Description	Company	Status					
			Idea	Business Case	Scoping	Specification	Agile Execution	Implementation
Data Converter	•A temporary, expedient solution to transform data into ADF to mitigate obsolescence; enables first step to full adoption	A						
		B						
		C						
Lab & Plant Automation	•Platform for the planning, execution, analysis & reporting of analytical chemistry leveraging the Framework •Includes IoT instrument integration; metadata repository/method management; workflow execution; ADF I/O integration with COTS and in-house software •Enables significant opportunities for automating data flow	D						
		E						
		F						
		G						
		H						
		I						
Taxonomies	•Leverage Allotrope taxonomies to provide metadata for enriched index •Lightweight universal viewer for any technique	H						
		J						
Data Lake	•Repository based on ADF/AT/APIs	H						
CRO Integration	•Convert at CRO to return raw & processed data back to company in ADF	K						



2016
- Framework used in production
- First Public release

2015
- Framework Development
- Integration at Members

2014
- Established feasibility through PoCs
- ADF design & due diligence

2013
- Initiated software development and evaluations

2012
Allotrope Foundation

INTERNATIONAL CONSORTIUM for INNOVATION & QUALITY in PHARMACEUTICAL DEVELOPMENT

Thank you!

Networking with Peers: upcoming workshops and meetings

- Sep 15, 2015 (Chicago, IL): Allotrope Partner Network F2F Workshop
- Sep 16, 2015 (Chicago, IL): Cross Industry Workshop

To join or get additional information, contact:

James Vergis, Ph.D.
Science Advisor | Drinker Biddle & Reath LLP
1-202-230-5439
James.Vergis@dbr.com
more.info@allotrope.org **www.allotrope.org**

What brought me to Uppsala ??

- Meeting Thomas Balzer at an IDMP conference in Berlin
- Allotrope Framework presented as a long-term option to establish a smooth dataflow
  - from analytical data creation in the lab
  - to analytical Data reporting e.g. for Group 4 Specified Substances
- Allotrope closing a gap in terms of missing analytical data standards in the CMC area
- Discussion:
  - Can we identify synergies between Allotrope Framework for analytical data management in the Lab and analytical information in a substance database

Specified Substance Group 4 Analytical Data

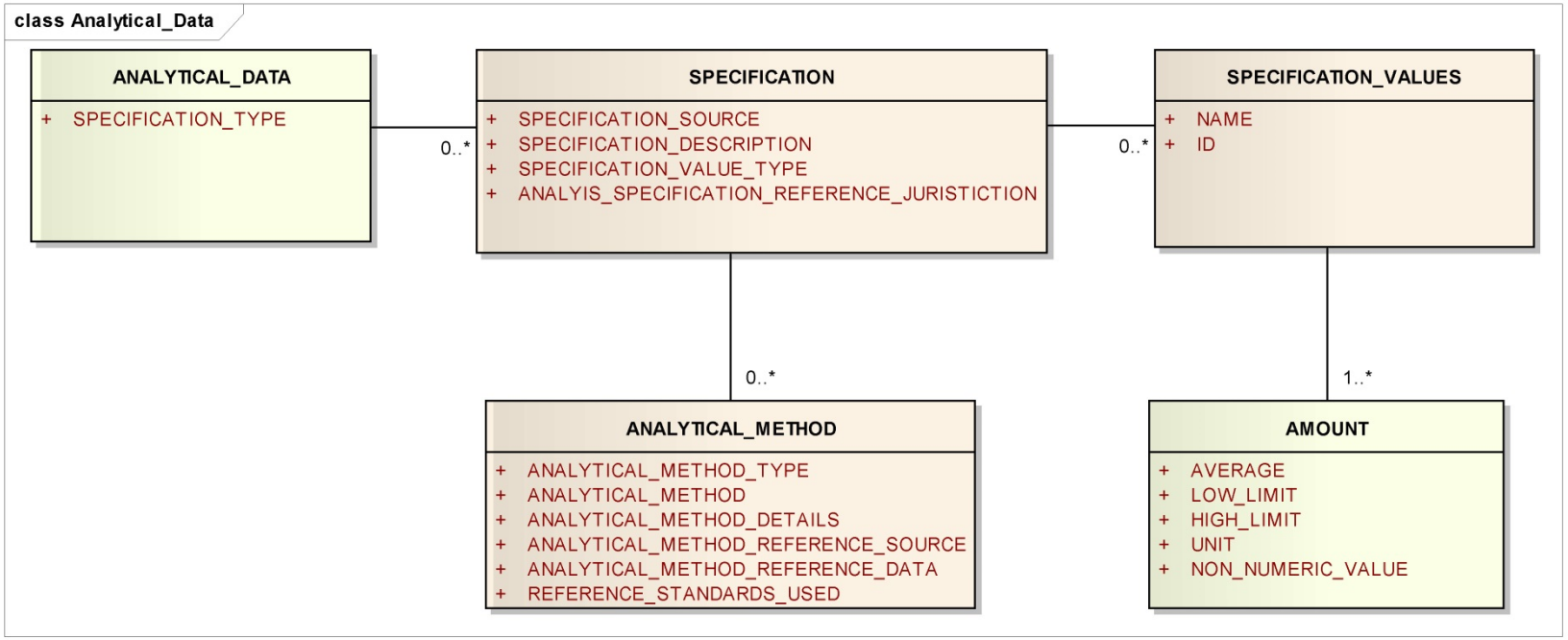
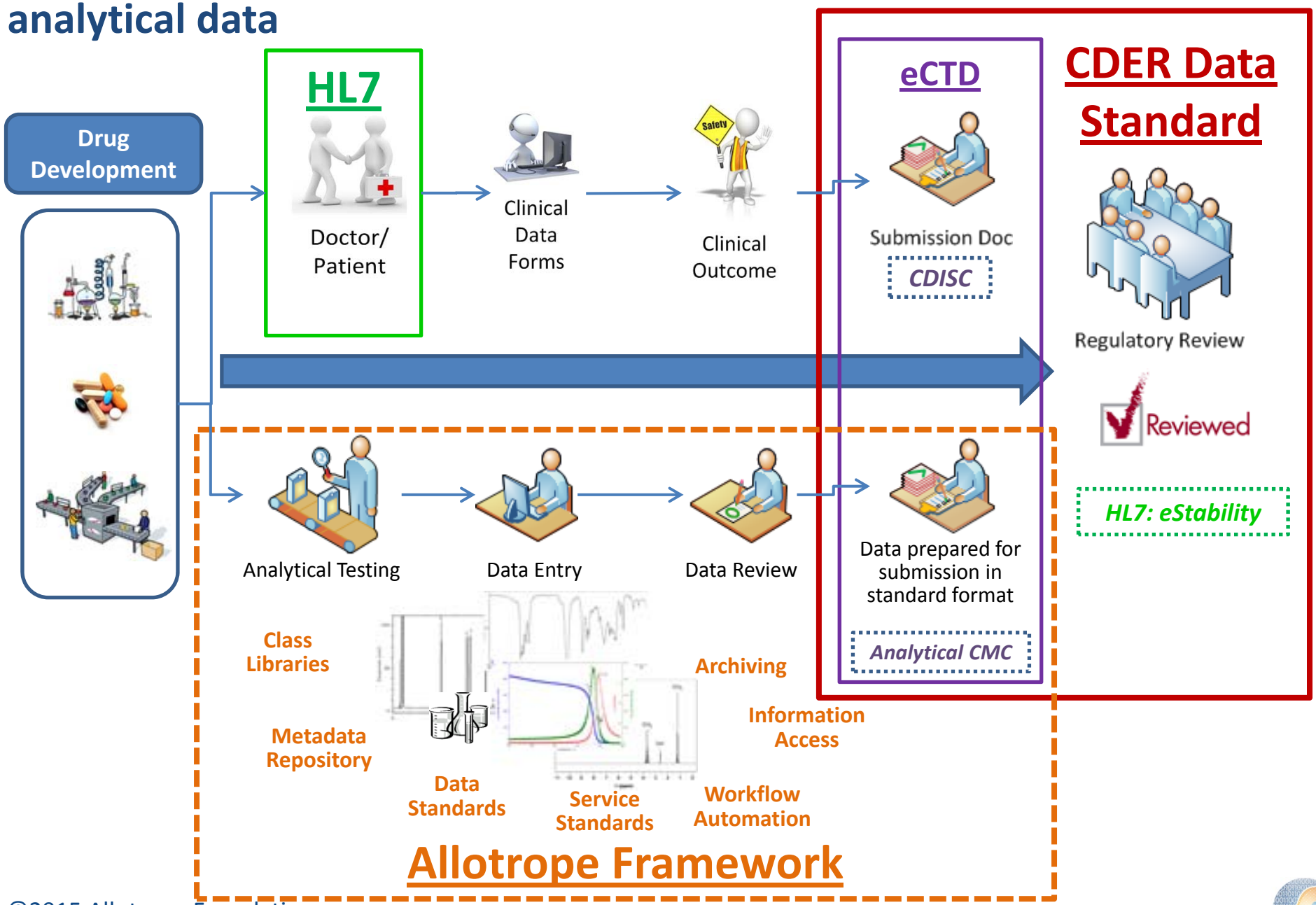


Figure 19 Information model for the analytical data class

These are Data that are in our:
LIMS (Laboratory Information Management System) and
ELN (electronic Lab Notebook) and CDS (Chromatographic Data systems) today
Raw Data from Instruments in many different mainly proprietary Data formats
Processed Data (results) with non-harmonized meta-data

Allotrope Framework addresses the gap in standards for CMC analytical data

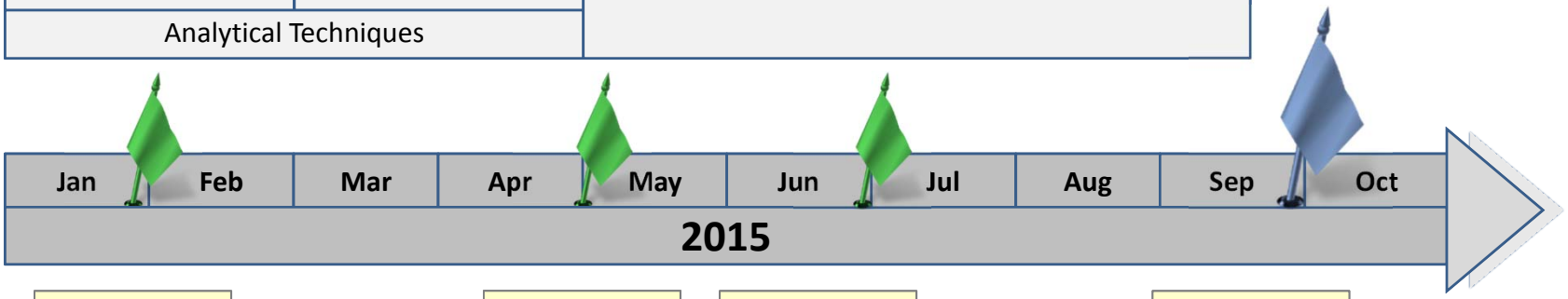


September, 2015 Release: Version 1.0

Draft ADF Format & API	ADF Format & API Version 1.0
Data Description API	Data Description API Version 1.0
Data Cube API	Data Cube API Version 1.0
Data Package API	Data Package API Version 1.0

Draft Analytical Data API	Analytical Data API Version 1.0

Draft Taxonomies		Taxonomies Version 1.0
Equipment	Process	Taxonomies Version 1.0
Material	Result	
Analytical Techniques		



Draft Data Description API

Alpha release to APN

Draft Analytical Data API

Release Version 1.0

ISO IDMP Implementation	CDER CBER	Implement International Organization for Standardization (ISO) Identification of Medicinal Products (IDMP) standards with reliable and robust repositories and processes to support efficient, consistent, and timely decision making in the regulation of medicinal product throughout the product development lifecycle.	ISO 11238 pilot was completed based on the Global Substance Registration System (GSRS) in April 2015 with positive outome. Project is moving to establish GSRS in production environment and Phase 1 to be completed FY2016 Q1.