

**BEST
PRACTICES
AWARDS**



WINNER

Bio-IT World '17
CONFERENCE & EXPO

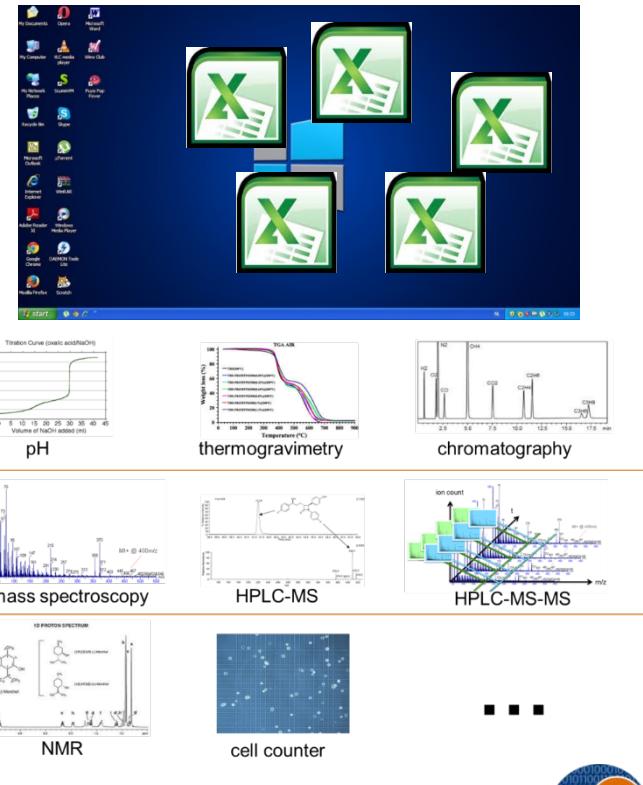
The Allotrope Framework 101

Allotrope Connect, November 6, 2018

Status quo in the laboratory

Many challenges for data capture, integration & sharing

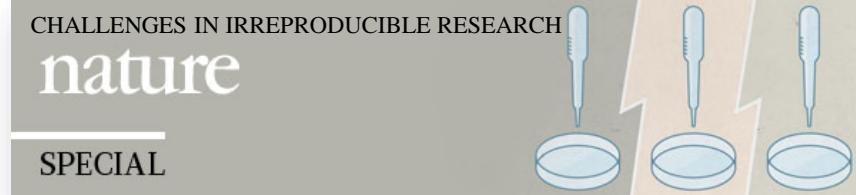
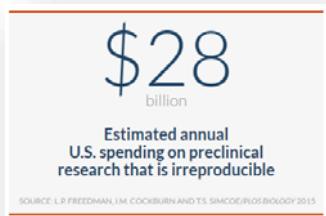
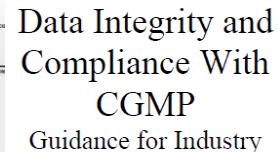
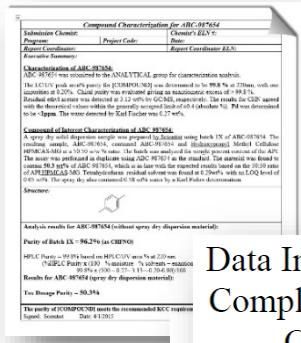
- Manual transcription of methods
- Incompatible instruments & software systems
- Manual data handling that challenges data integrity
- Legacy architectures are brittle & rigid
- Knowledge & context resides in people's heads
- Silos of data, context and meaning



From Eric Little, OSTHUS



**Too much paper + manual tasks =
delayed medicine to patients & erosion of public confidence in data integrity**



©2018 Allotrope Foundation



Better Scientific Reproducibility



Reduce Manual Effort & Paper



Increase Data Integrity, Context, Quality



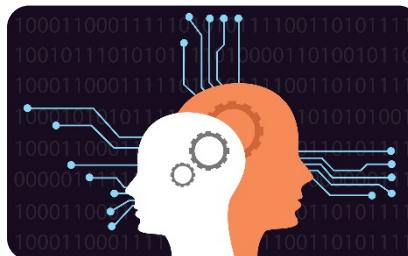
Streamlined Access, Sharing, Integration



Simplified IT



Foundation for Data Science



Consolidate Requirements Lower Innovation Barrier



Reduced Manual Effort & Paper



Better Scientific Reproducibility



Streamlined Access, Sharing, Integration



Simplified IT



Increased Data Integrity, Context, Quality



Foundation for Data Science



Consolidated Requirements Lower Innovation Barrier



Standardized Data & Metadata

How do we deliver this?

The Allotrope Framework

A standardised semantic model for data & metadata.

Released Dec 2017



Released Dec 2018

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

Dec 2018

Released July 2017

Allotrope Data Format

Data Description
Semantic Model

Data Cubes
Universal Data Container

Data Package
Virtual file system

API

```
class Directory
    private Directory[] mySubs;
    private File[] myFiles;

    public Directory()
    {
        int numSubs = (int)0;
        mySubs = new Director[0];
        int numFiles = (int)0;
        myFiles = new File[0];
    }

    for(int i = 0; i < mySubs.length; i++)
        mySubs[i] = new Directory();

    for(int i = 0; i < myFiles.length; i++)
        myFiles[i] = new File();
    }
```

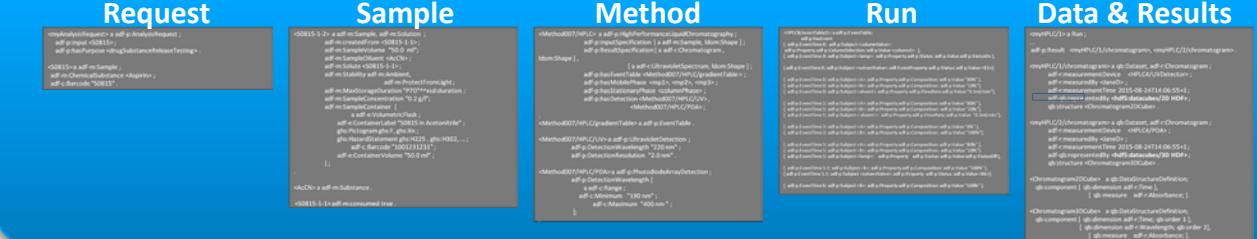
A high-performance binary data format. Instrument, vendor, platform agnostic.

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



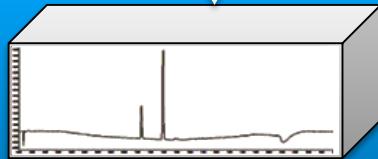
Allotrope Data Format Example

Data Description

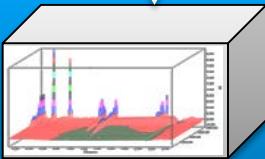


Data Cubes

Chromatogram: 2D



Chromatogram: 3D



Data Package



Platform Independent File Format

©2018 Allotrope Foundation

Descriptive metadata about

- Method, instrument, sample, process, result, etc.
 - Data Cube, Data Package contents
 - Provenance, audit trail, data models

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

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

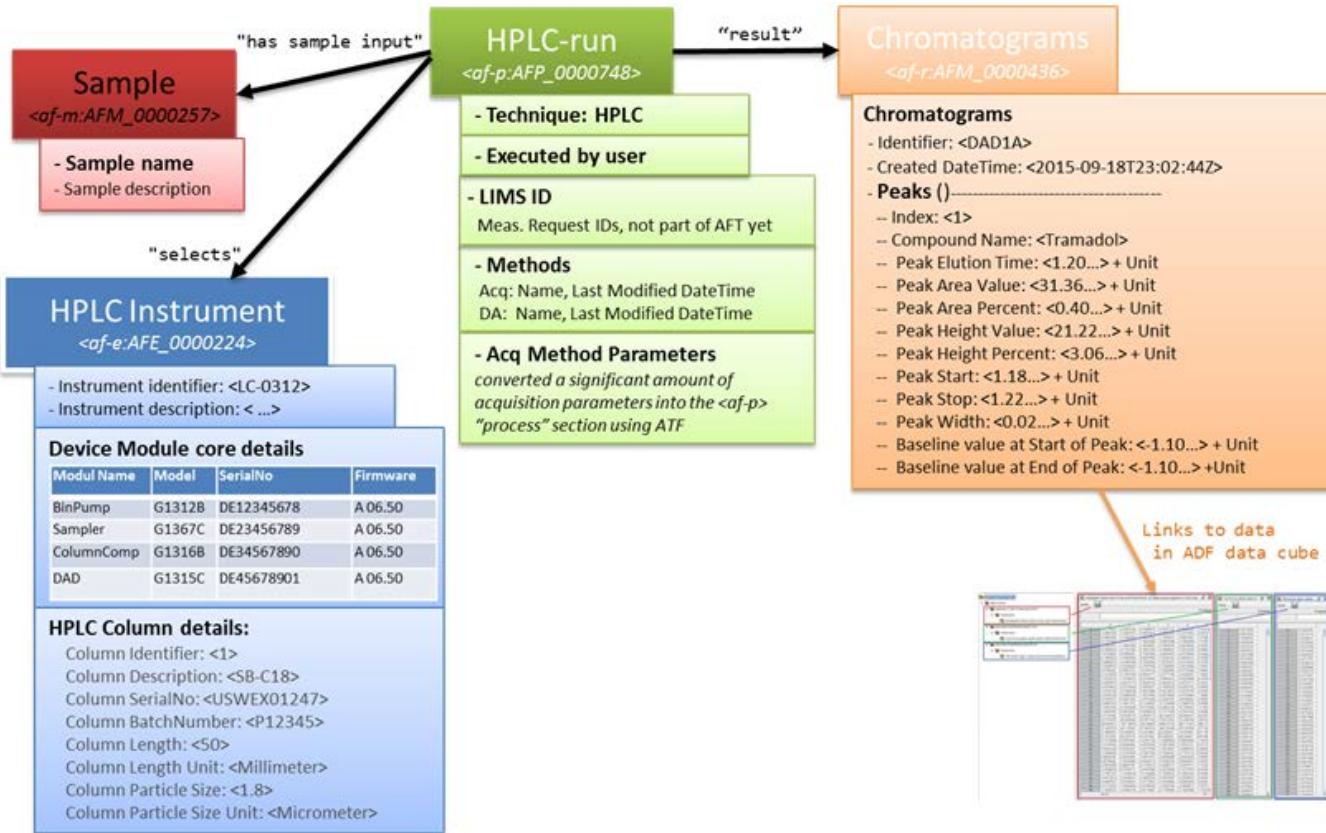


Allotrope Data Format - Development in 2018

- Version 1.4.3 RC released for testing
- Audit Trail for Data Package
 - Rename file
 - Replace file
- Upgrade files from older version (1.3.1+)
- 5 Bug fixes



Example: ChemStation Metadata Content



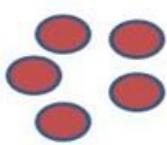
Allotrope
Foundation
TaxonomyAllotrope
Foundation
OntologyAllotrope Domain
Ontologies

Terminologies

References to
External Ontologies
(Curated)

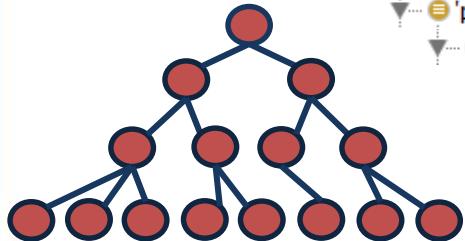
Ontologies

Terminology



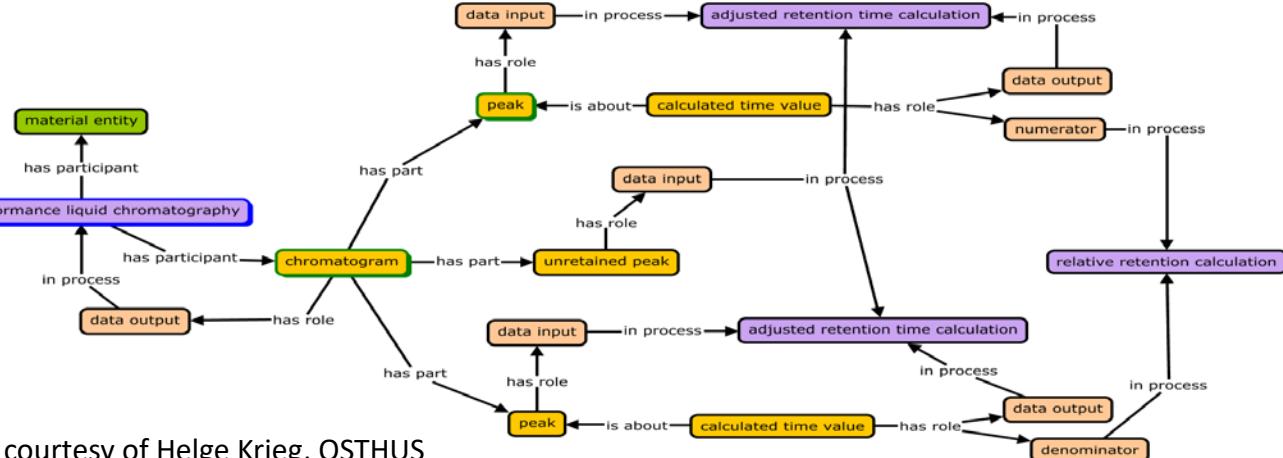
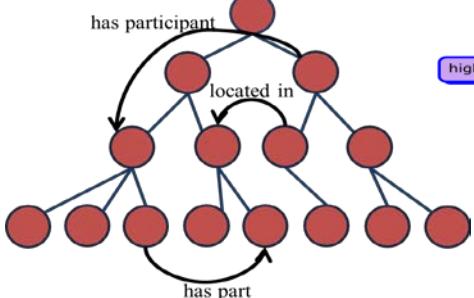
chromatography
column chromatography
liquid chromatography
chromatogram
peak
spectrum
ultraviolet spectrum

Taxonomy



occurred ≡ TemporalEntity ≡ AFC_0000079
 process ≡ AFP_0001617
 'planned process'
 'material processing' ≡ AFP_0003275
 'separation method'
 chromatography
 'column chromatography'
 'liquid chromatography'
 'liquid-solid chromatography'
 'high-performance liquid chromatography'

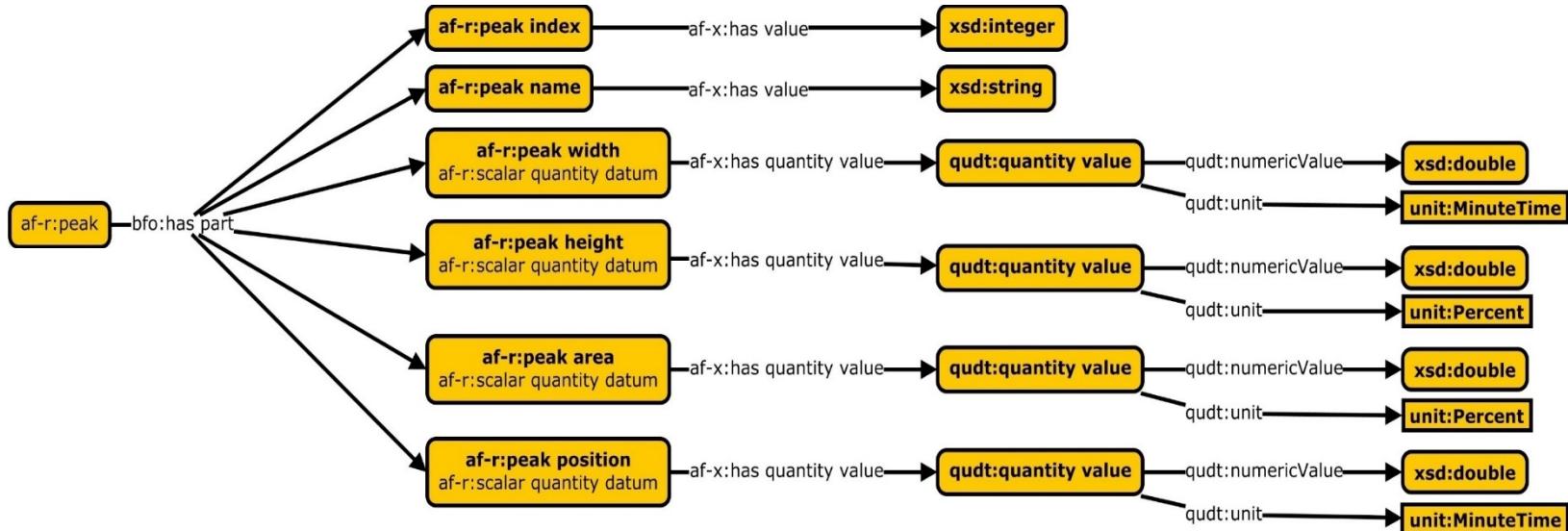
Ontology



Model courtesy of Helge Krieg, OSTHUS

Data Models: Example- HPLC

- Ontologies provide an unconstrained vocabulary we can use to describe things (instances) in our open world and give them a meaning (= **what** it is)
- Data structures (schemas, templates) describe **how** to use the ontologies for a given purpose in a standardized (reproducible, predictable, verifiable) way
- Shapes Constraint Language (SHACL, expressed as RDF) is a WC3 standard to do this, used for Allotrope ‘Data Models’



The Allotrope Data Models - Development in 2018

Design Pattern Catalog

A set of reusable “lego blocks” that standardizes the modelling approach

= The Modelling Framework

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

Data Models

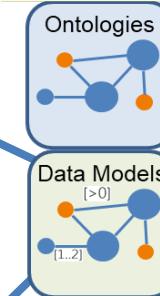
A ever-growing list of data models for various techniques:

LC, GC, DSC, MS, BALANCE, ...

= Applications of the Modelling Framework

A standardised semantic model for data & metadata.

Released Dec 2017



Released Dec 2018

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

Released July 2017

Allotrope Data Format

Data Description
Semantic Model

Data Cubes
Universal Data Container

Data Package
Virtual file system

API

```
private Directory myDir;
private Directory[] mySubs;
private File[] myFiles;
```

```
public Directory()
{
    int numSubs = (int) (Math.random() * 10);
    mySubs = new Directory[numSubs];
    int numFiles = (int) (Math.random() * 10);
    myFiles = new File[numFiles];
}

for(int i = 0; i < numSubs; i++)
    mySubs[i] = new Directory();
for(int i = 0; i < numFiles; i++)
    myFiles[i] = new File("file" + i);
```

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



Design Pattern Catalog



Authors:

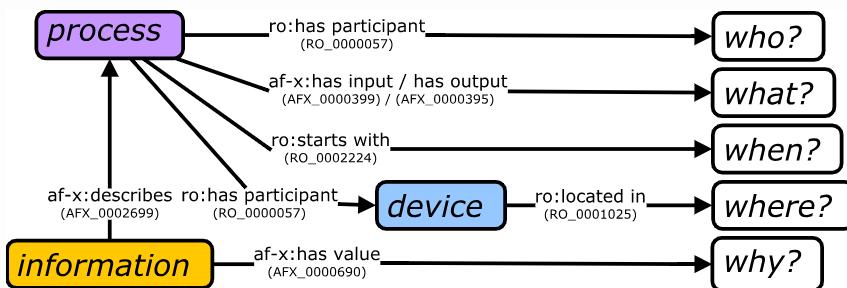
Thomas Weber, OSTHUS GmbH
Helge Krieg, OSTHUS GmbH

[Collapse all](#)



Semantic pattern of audit trail

Semantic pattern of *audit trail* is a set of reusable structures for description of audit trail of processes answering core questions of *who?*, *what?*, *when?*, *where?*, *why?*. It is an extension of [semantic pattern of datum](#).



▼ Pattern Name and Classification

Name: Semantic pattern of audit trail

Classification: Semantic Pattern

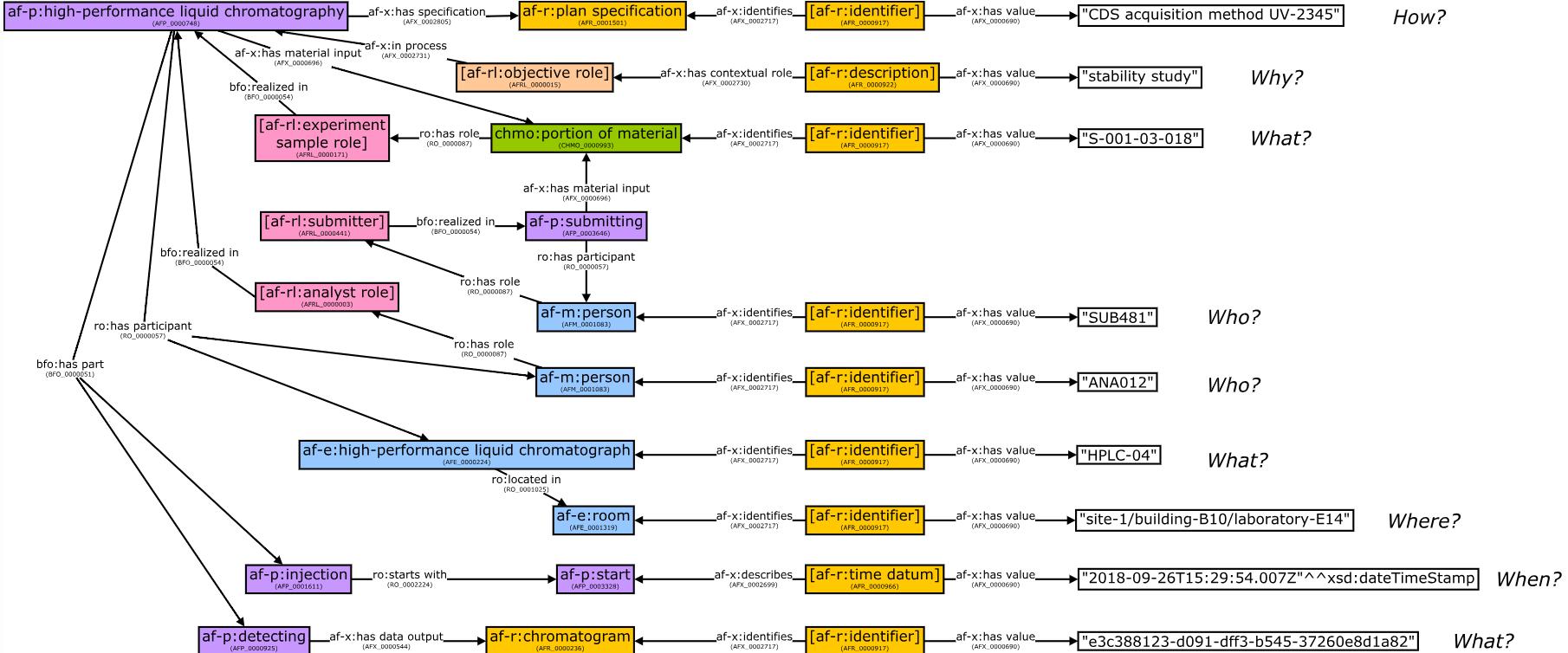
▼ Also Known As

- 5W pattern
- core questions pattern

Known Uses

- follows a defined specification. → how?

[Collapse all](#)



▼ Related Patterns

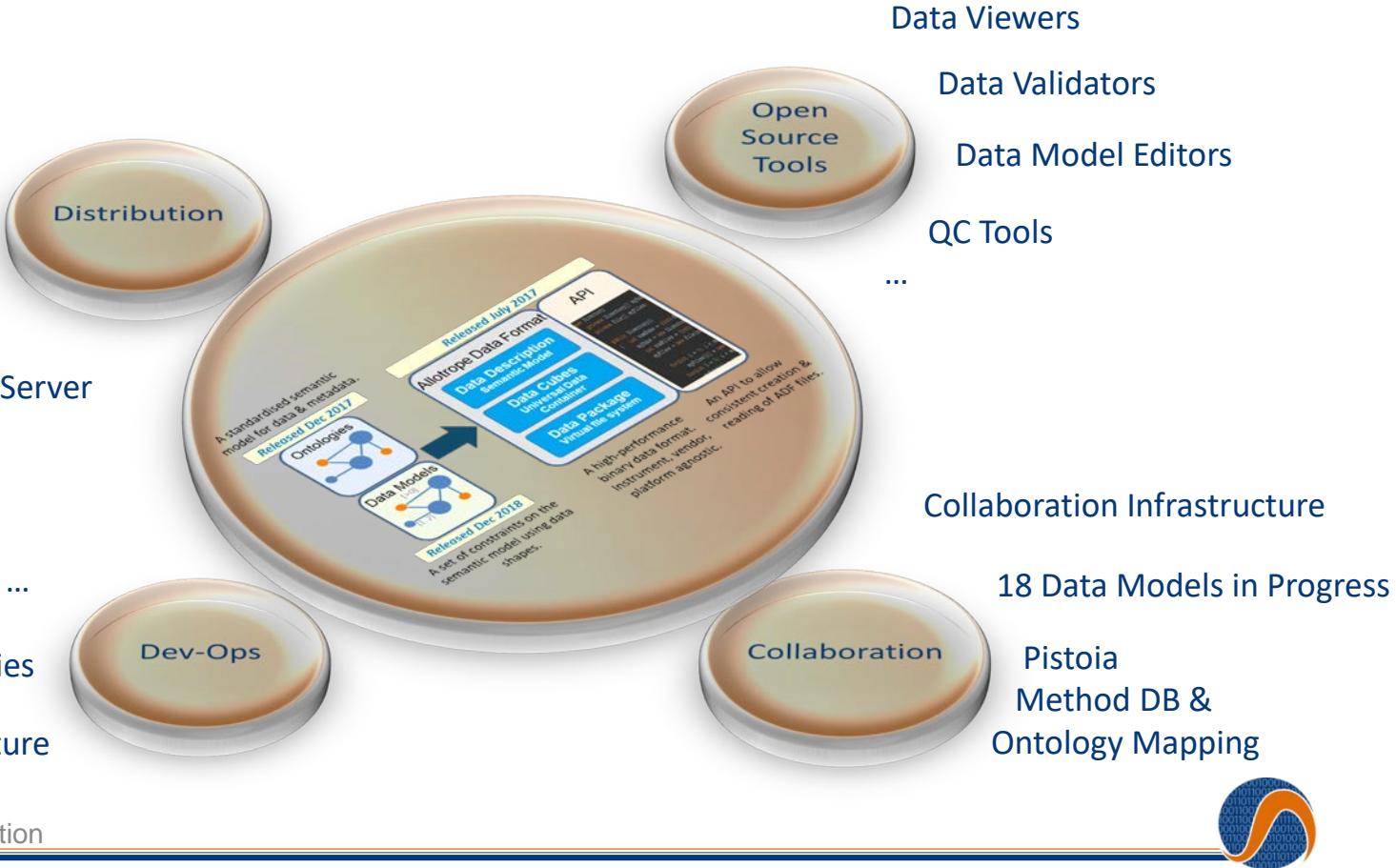
The Allotrope Ecosystem

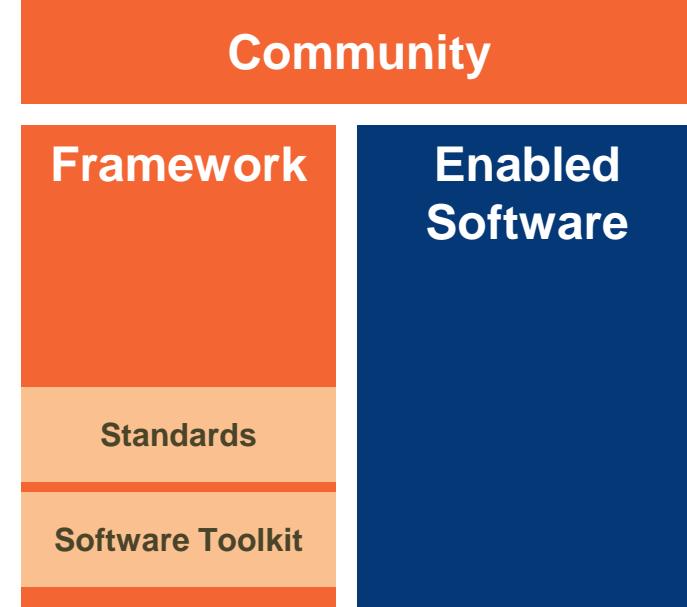
Public Repositories?
e.g. <http://obi-ontology.org>

Allotrope Repository
<https://allotrope.ifrog.io>

Allotrope Documentation Server
<https://docs.allotrope.org>

Automated QC for Ontologies
Automation Infrastructure





The Allotrope Community



Allotrope
Foundation

abbvie

AMGEN®

Baxter



OSTHUS®
success with R&D



Biogen

BAYER
BAYER

Bristol-Myers Squibb

Genentech
A Member of the Roche Group

gsk

Lilly

MERCK
INVENTING FOR LIFE

novo nordisk®

Pfizer

Drinker Biddle®



COGNITIVECHEM

HDF
The HDF Group

LABLICATE
peak efficiency

Malvern Panalytical

DEXSTR

Abbott

PERSISTENT

SHIMADZU
Excellence in Science

iuta

MESTRELAB RESEARCH
Chemistry Software Solutions

SYNTHACE

GE

Cognizant

PerkinElmer®
For the Better

BIOVIA

GTY

ZONTAL...
KNOWLEDGE CONNECTED

HALO
DIGITAL

idbs

Riffyn

BIO-RAD

Cytobank

Science & Technology
Facilities Council

sartorius

accenture

PANGAEA
Bringing it together to make it happen

Waters
THE SCIENCE OF
WHAT'S POSSIBLE.®

LAMSADE
UMR CNRS 7243

paradigm4

ELEMENTAL
MACHINES

NPL

SCIEX

UNCHAINED
LABS

BRUKER

University of
Strathclyde
Glasgow

ACD/Labs

ThermoFisher
SCIENTIFIC

Brigham Young University • BSSN Software • Elemental Machines • Erasmus MC • Fraunhofer IPA • L7 Informatics
Mettler Toledo • NIST • SciBite • Stanford University • University of Illinois at Chicago • University of Southampton

