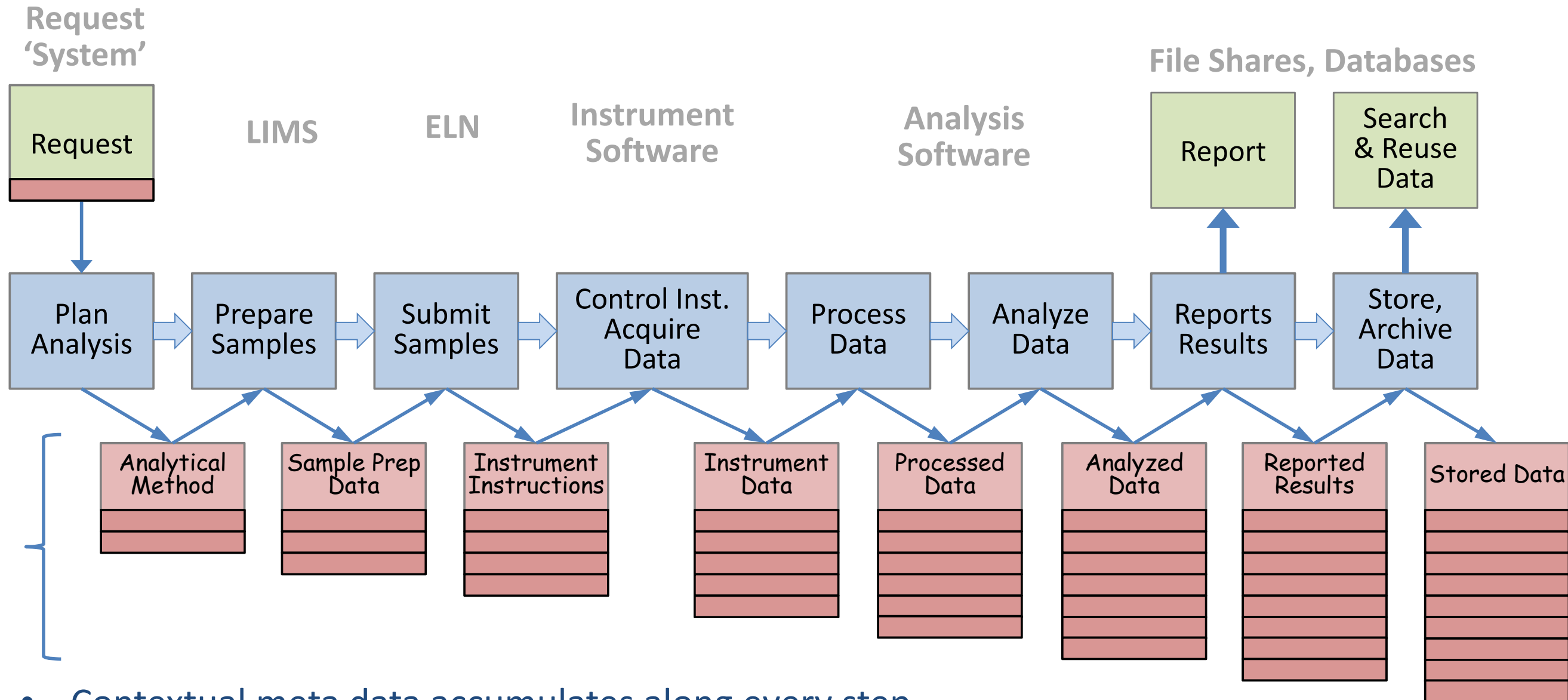


**Allotrope Foundation:
A Framework for Simplified Data Integrity,
Reproducibility and Data Management
through Improved Data Modeling With
Semantic Technologies**

Dana Vanderwall
Bristol Myers-Squibb & Allotrope Foundation

Fundamentals of a measurement workflow



- Contextual meta data accumulates along every step
- Distributed across multiple systems and records

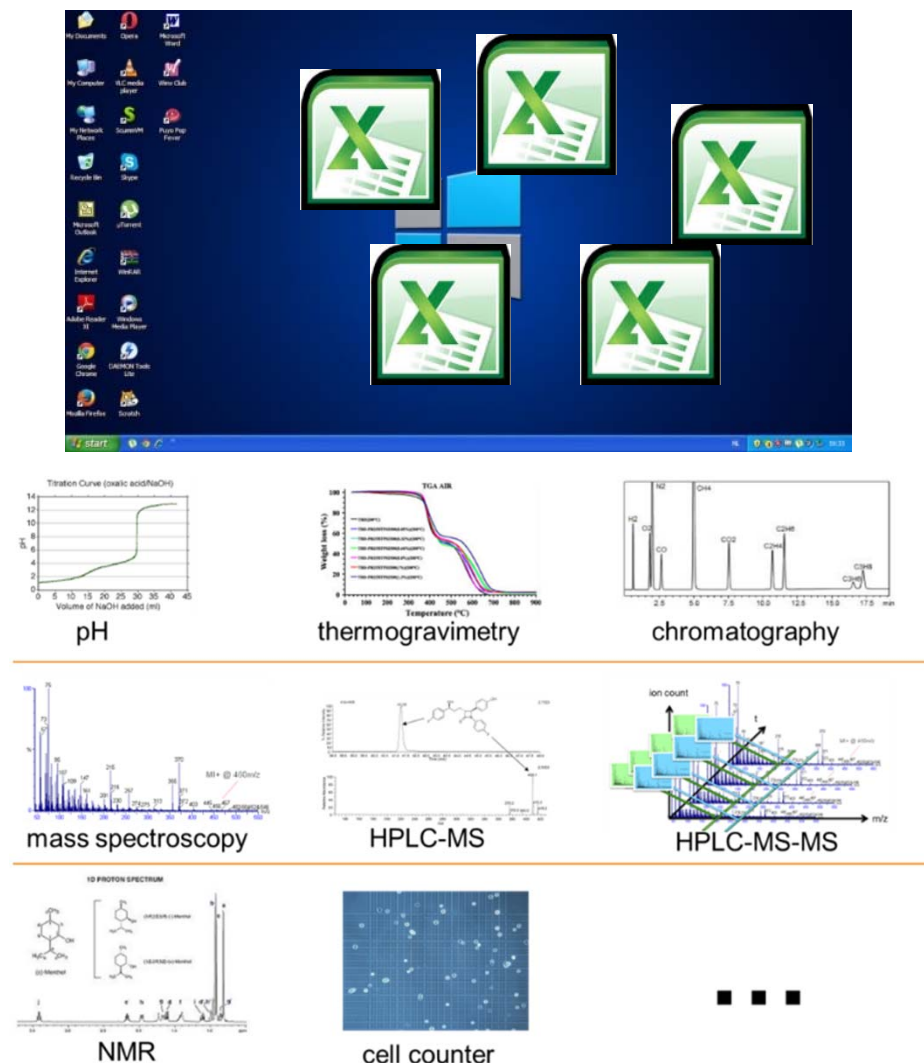


Status quo in the laboratory

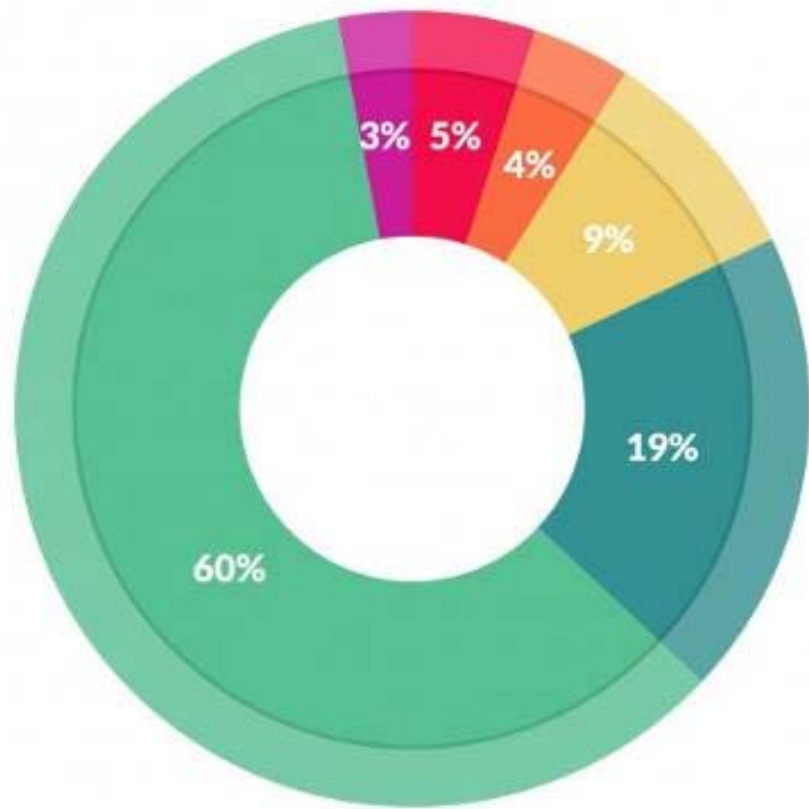
Many challenges for data capture, integration & sharing

- 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



Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says



What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets: 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

Gil Press, www.forbes.com, March 23, 2016; based on ClowdFlower survey of data scientists



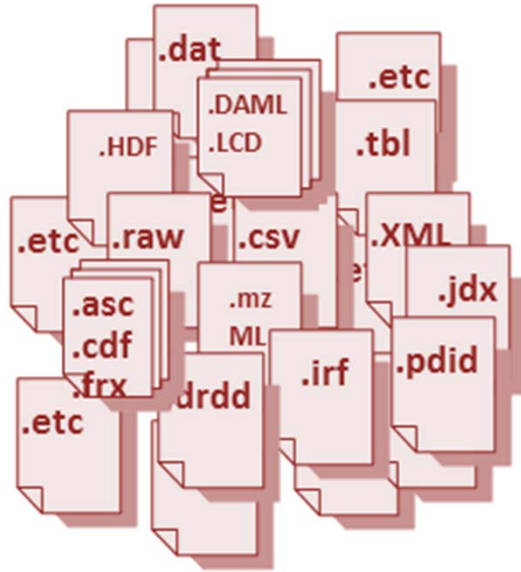
Data-Driven or Data Challenged Industry?

Data Problem **+** **Big Data** **=** **Big Data Problem**

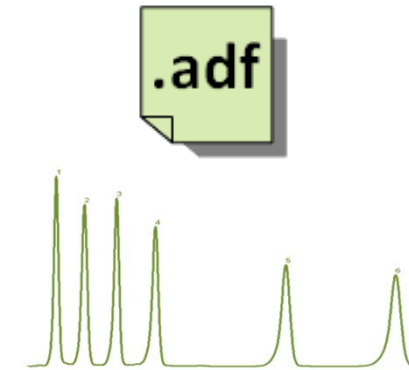


Addressing the root causes

Vendor/instrument-Specific Formats



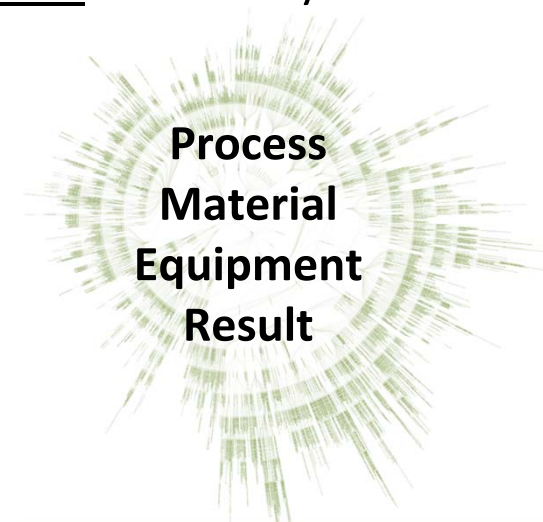
Data in Standard Format



Paper-based and unstructured text for methods, regulations, recipes, observations, etc

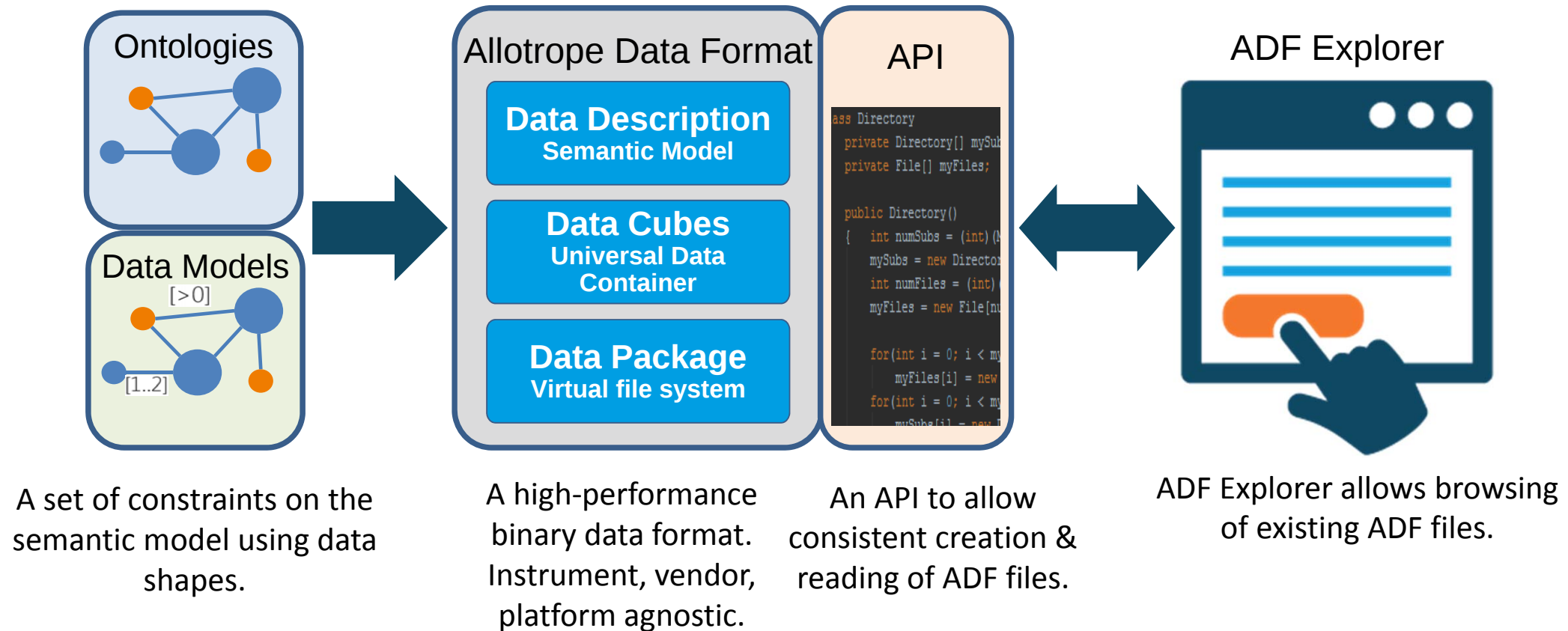


A Standard vocabulary & metadata structure

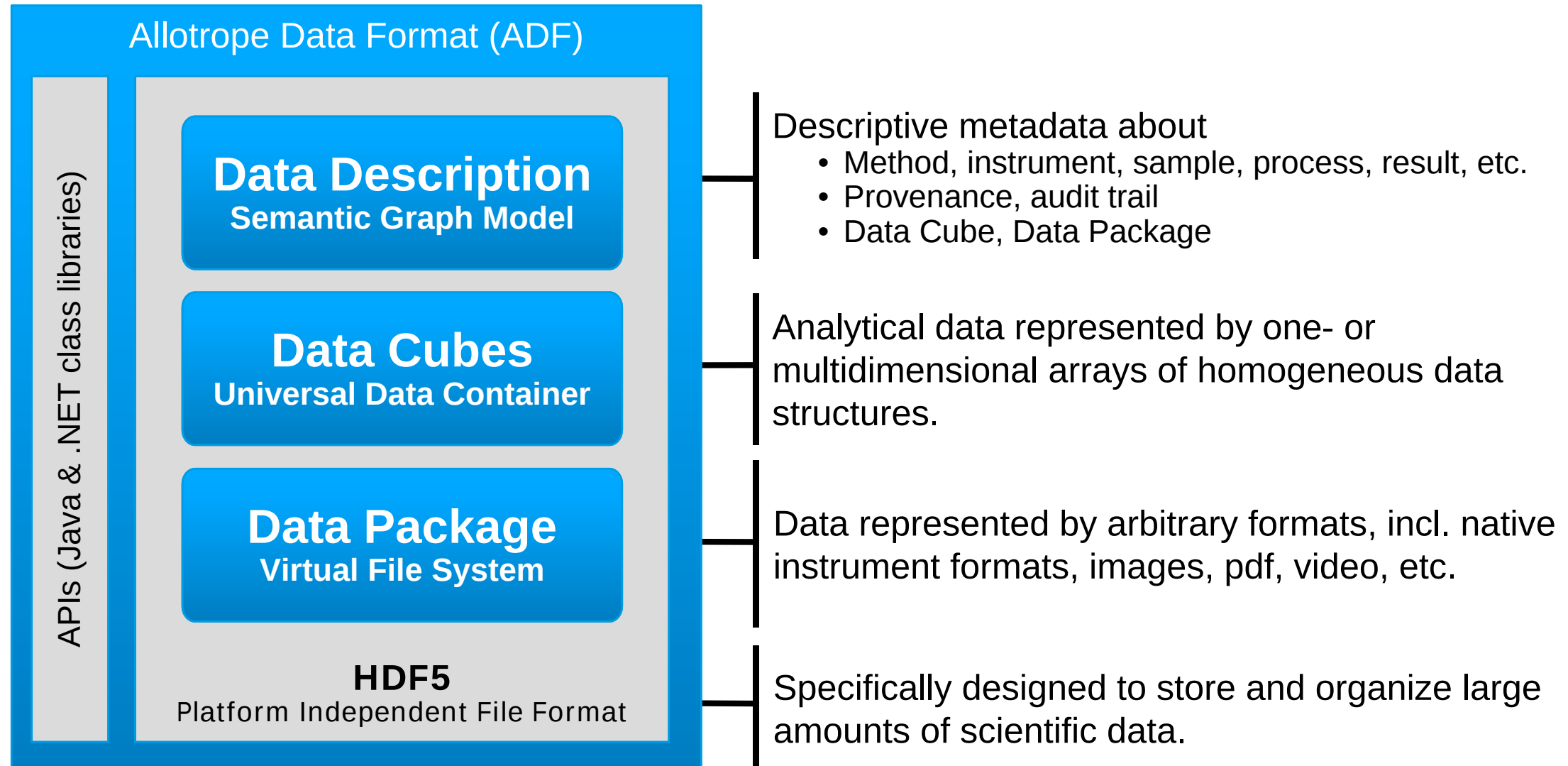


The Allotrope Framework

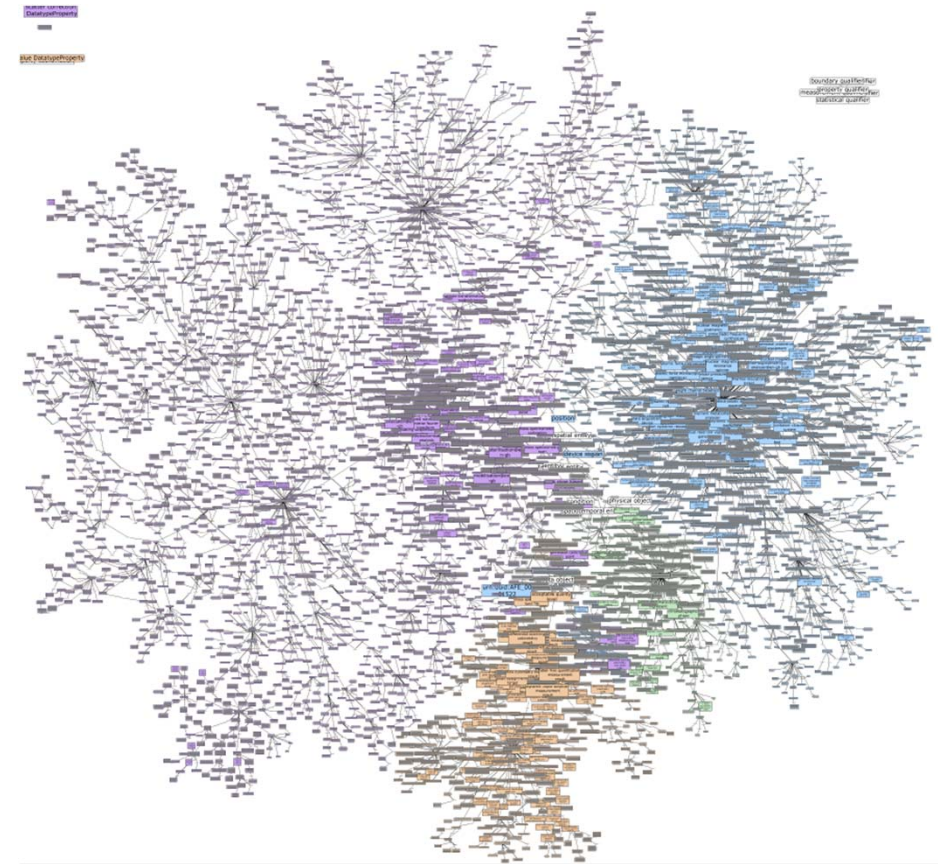
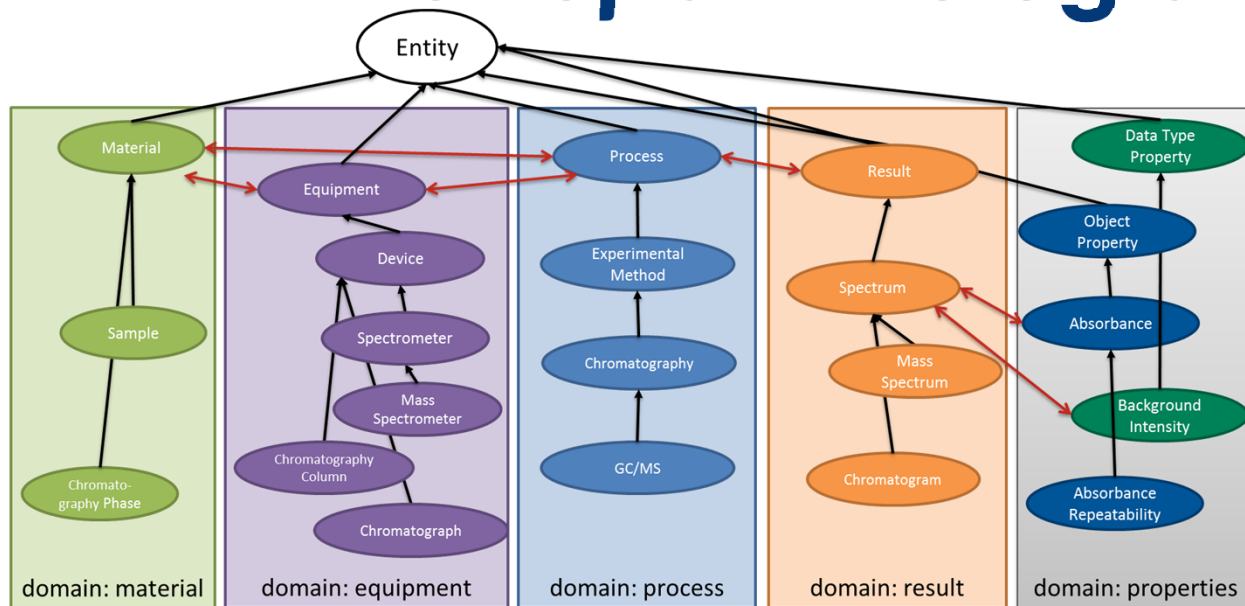
A standardised semantic model for data & metadata.



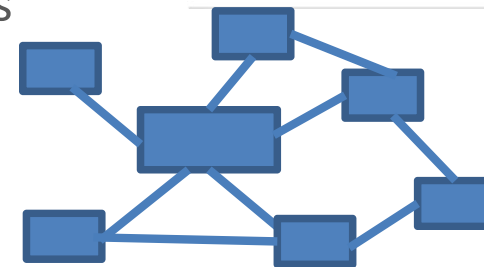
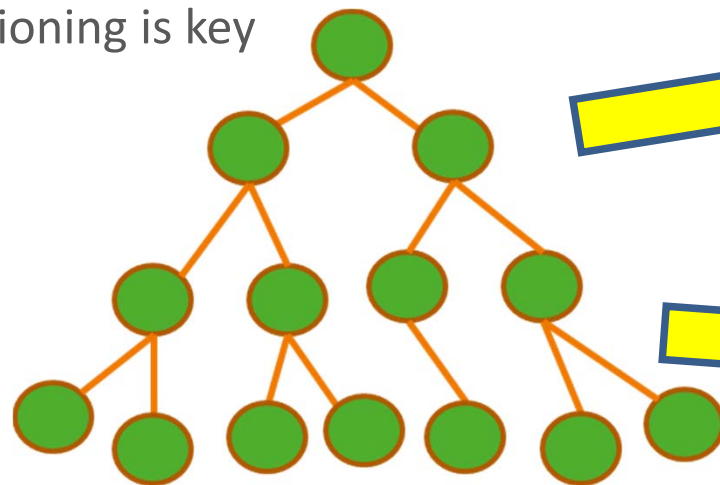
Allotrope Data Format (ADF)



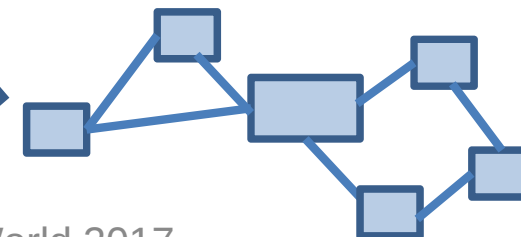
Allotrope Ontologies



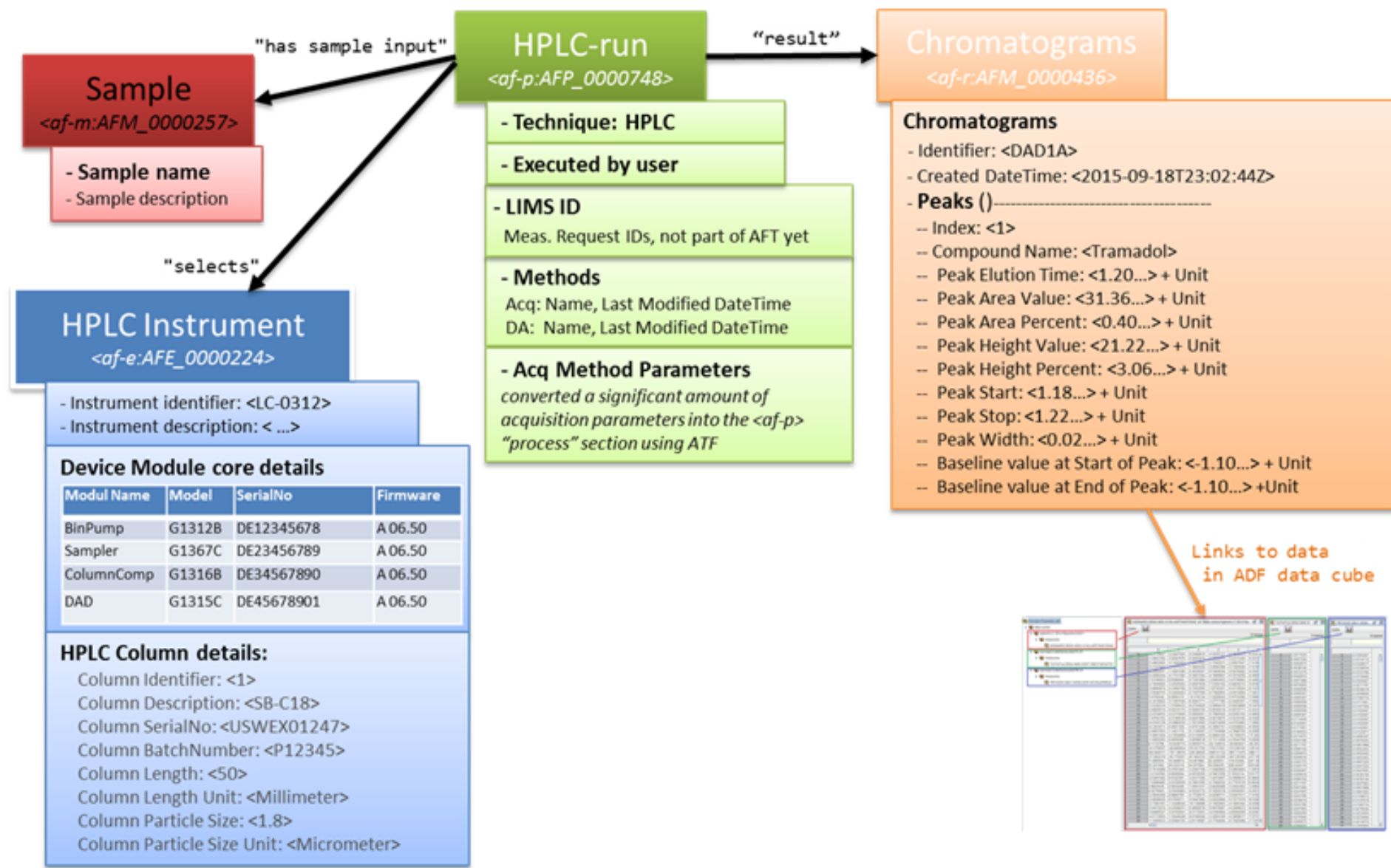
- Terms and definitions are contained in the taxonomy
 - What it is and where it is (URI)
- The taxonomy contains the smallest “building block” items
 - Modular Partitioning is key



Parts are reconstructed in the ontologies to leverage more complex relations



Example: ChemStation Metadata Content



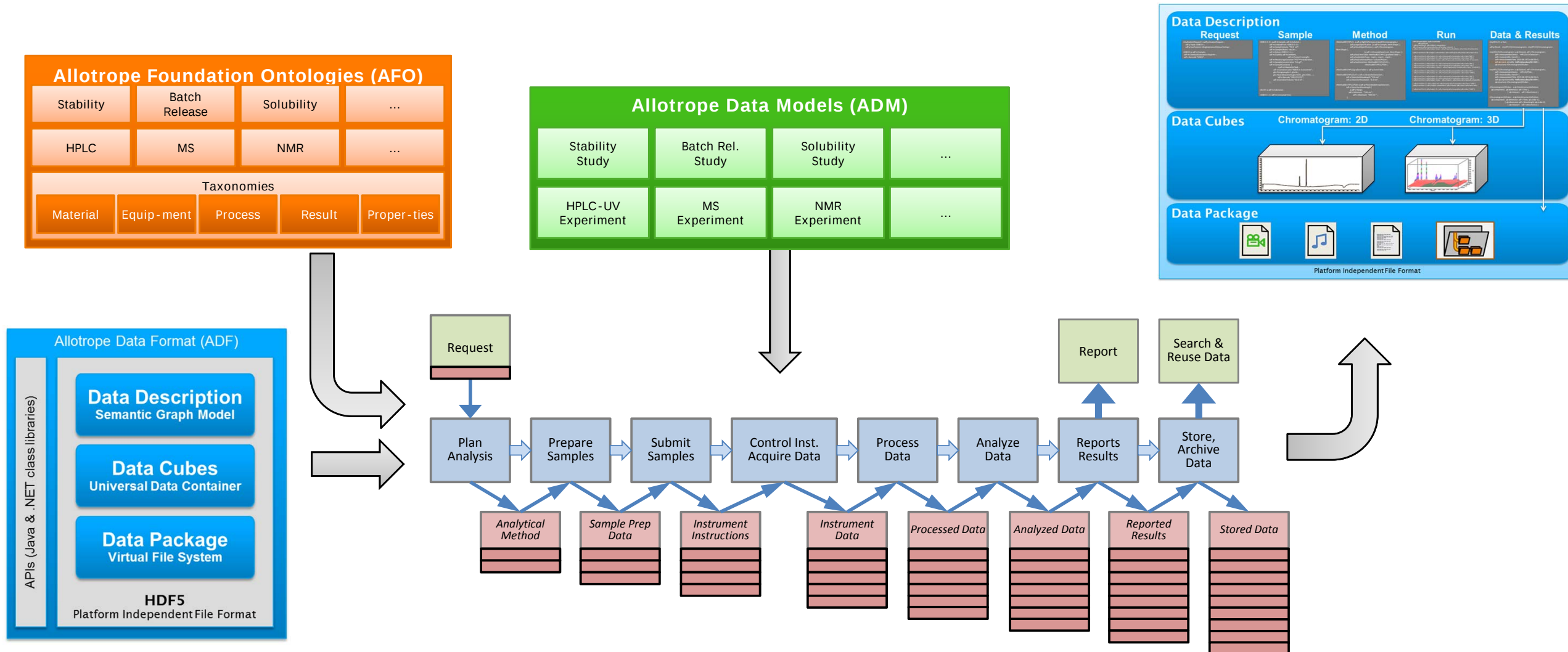
The Cmap represents the LC-UV model & datashape 2017/03/06

maximization of cardinalities (e.g. example of "has part")

- has part? → The relation is **applied**. Cardinality is "has 0". This is the default value.
- has part 1? → The relation is **mandatory**. There must be exactly 1 related object. Cardinality is "exactly 1".
- has part >1? → The relation is **mandatory**. There must be 2 or more related objects. Cardinality is "min 2".
- has part 1..4? → The relation is **mandatory**. There must be 1 to 4 related objects. Cardinality is "min 1 max 4".

-
- The mind map is organized as follows:
- HPLC run**
 - liquid chromatograph**
 - lab system
 - lab computer
 - method editor
 - data acquisition
 - report generator
 - method editor
 - data acquisition
 - report generator
 - valve
 - method editor
 - data acquisition
 - report generator
 - pump
 - method editor
 - data acquisition
 - report generator
 - autosampler
 - method editor
 - data acquisition
 - report generator
 - column oven
 - method editor
 - data acquisition
 - report generator
 - column
 - method editor
 - data acquisition
 - report generator
 - detector
 - method editor
 - data acquisition
 - report generator
 - channel
 - method editor
 - data acquisition
 - report generator
 - sequence of injections (flow injection)**
 - injection
 - method editor
 - data acquisition
 - report generator
 - separation
 - method editor
 - data acquisition
 - report generator
 - mixing
 - method editor
 - data acquisition
 - report generator
 - rinsing
 - method editor
 - data acquisition
 - report generator
 - sampling
 - method editor
 - data acquisition
 - report generator
 - detection
 - method editor
 - data acquisition
 - report generator
 - Full UV spectrum**
 - method editor
 - data acquisition
 - report generator
 - chromatogram**
 - method editor
 - data acquisition
 - report generator
 - peak list**
 - method editor
 - data acquisition
 - report generator
 - peak**
 - method editor
 - data acquisition
 - report generator

A Foundation for Interoperability & Next Generation Analytics

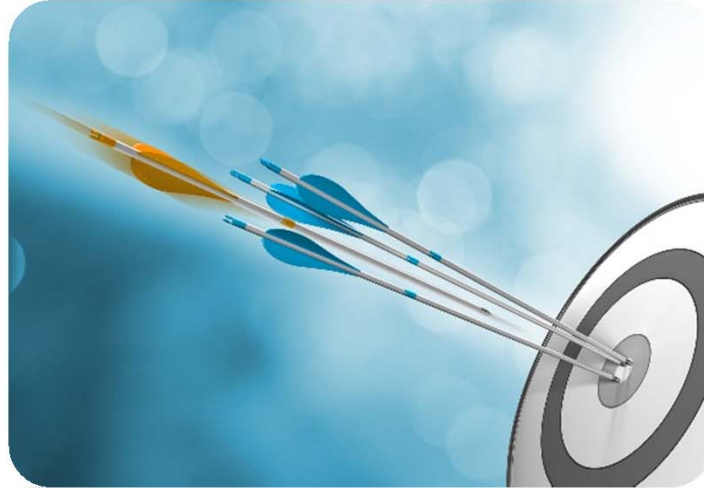


What can we achieve by rethinking how we collect and care for our scientific data?

Reduced Manual Effort & Paper



Better Scientific Reproducibility



Increased Data Integrity,
Context, Quality



Streamlined Access,
Sharing, Integration



Consolidated Requirements
Lower Innovation Barrier



Simplified IT



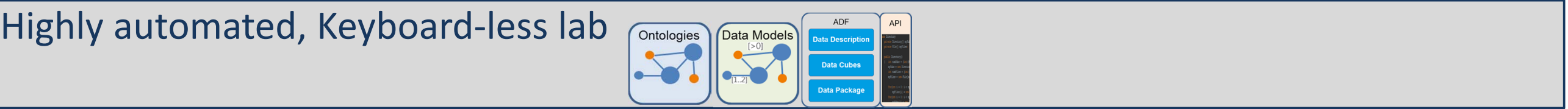
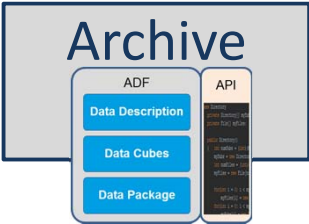
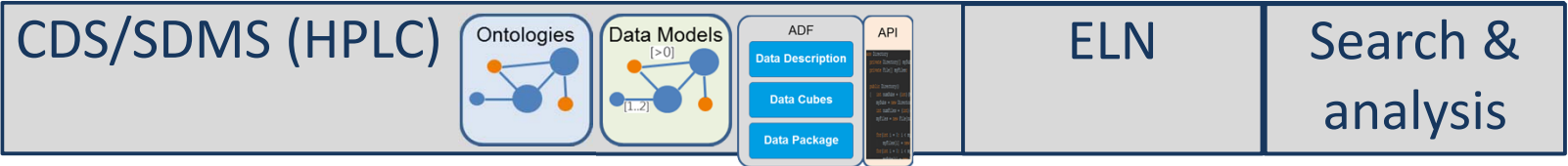
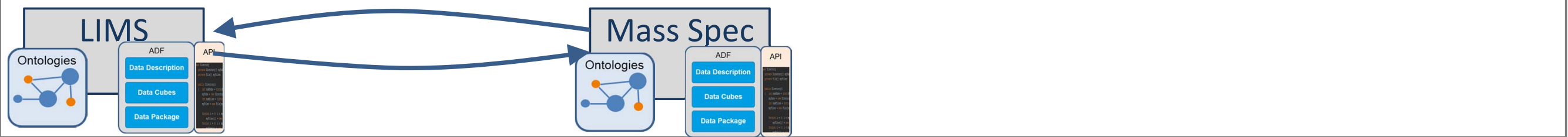
Foundation for Data
Science



Allotrope Foundation, BioIT World 2017



Examples of Work in Flight @ Allotrope Member Companies



Release Roadmap

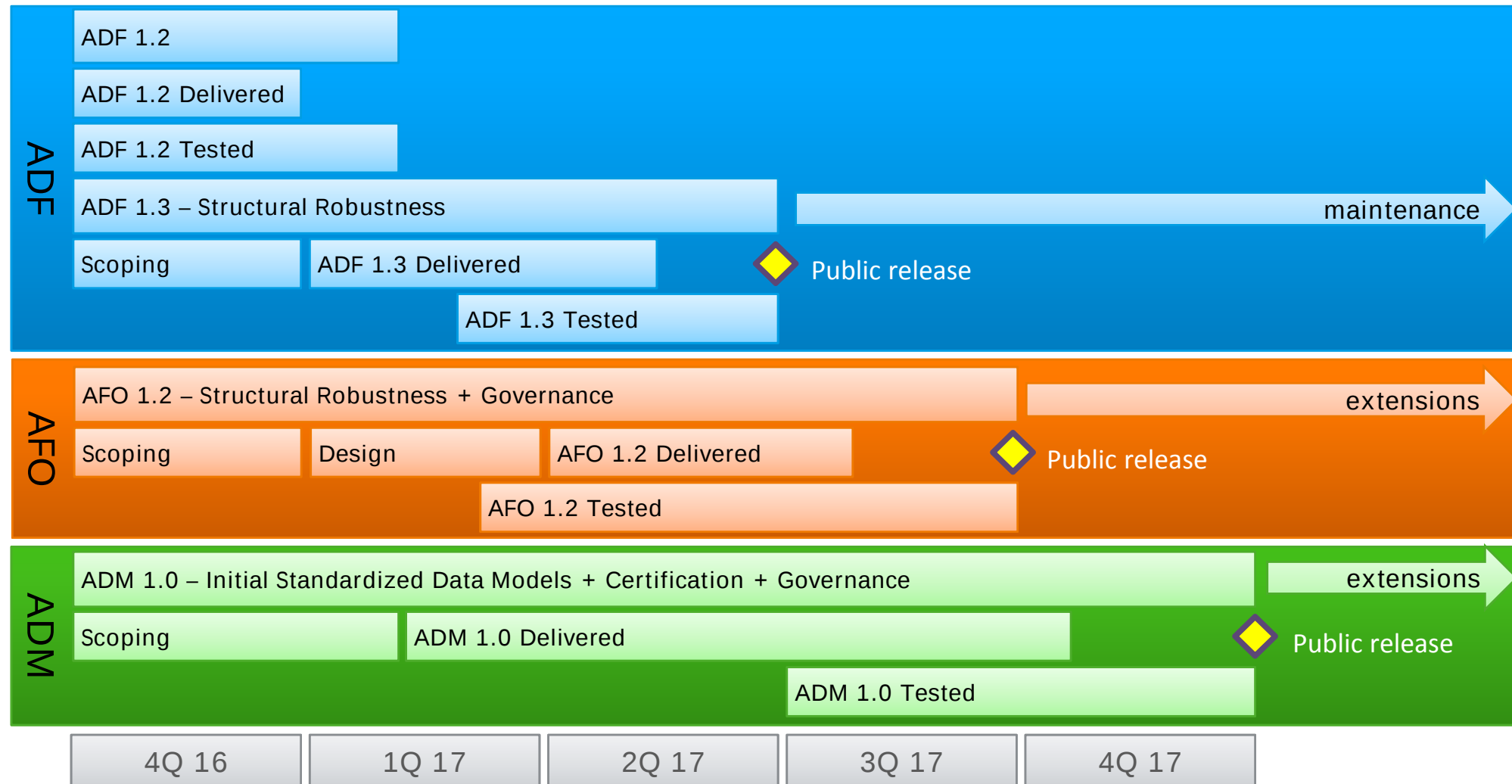
Version 1.0 released Sept 2015

ADF Allotrope Data Format

AFO Allotrope Foundation Ontologies

ADM Allotrope Data Models

- AFO & ADM releases in 2017 complete for HPLC-UV via Allotrope
- Other techniques likely in 2017: NMR, MS
- 'ADF Explorer' publically available July 2017



Working group examples

HPLC-UV WG

ACD Labs	Lilly
Agilent	Merck
Amgen	Pfizer
Biovia	Shimadzu
Genentech	Waters
GSK	Zifo

NMR WG

ACD Labs	Bruker
Baxter	GSK
Boehringer	Merck & Co.
Ingelheim	MestreLab
BMS	Pfizer

Governance Team

Agilent	Mettler Toledo
Amgen	Pfizer
Bayer	Riffyn
IDBS	Thermo Fisher
Merck	Waters

Mass Spec WG

AB Sciex	Merck
Agilent	Thermo Fisher
Bayer	Waters
BMS	

ADF WG

Amgen	Agilent
Bayer	Shimadzu
GSK	Thermo Fisher
Merck	Waters

Certification & Support Req. WG

Agilent	Merck
Baxter	Mettler Toledo
BMS	Persistent
IDBS	Thermo Fisher





AbbVie
Amgen
Baxter
Bayer
Biogen

Boehringer Ingelheim
Bristol-Myers Squibb
Eli Lilly
Genentech/Roche
GlaxoSmithKline

Merck & Co.
Novo Nordisk
Pfizer



Drinker Biddle®
Secretariat



BSSN Software • Cognizant • Erasmus MC • Fraunhofer IPA • Halo-Digital • The HDF Group LabAnswer
• LabWare • Mettler Toledo • Stanford University • University of Illinois at Chicago • University of
Southampton



Join Allotrope Foundation!

Influence direction
of development



Be part of an expanding
community of experts



Receive support & training



Benefit from shared Investment



Align internal strategy with
the future of data



Ensure sustainability & adoption



Allotrope Foundation, BioIT World 2017



Thank you!

- **2017 “Allotrope Connect” Workshops**
 - March 29-30, 2017 @Pfizer, Groton, CT
 - June 21-22, 2017 @GSK, WARE, UK
 - September 20-21, 2017 @Merck, New Jersey

More details and contact info: allotrope.org

Allotrope Foundation- Scientific data you can trust, at your fingertips.

Allotrope Foundation, BioIT World 2017

