Allotrope Framework

Revolutionizing the way we acquire, share and gain insights from scientific data, through a community and framework for standardization & linked data

Vincent Antonucci
Global Regulatory CMC, Merck & Co., Inc
on behalf of the Allotrope Foundation
October 2017

Public

Allotrope Framework: From concept to reality

2012

Allotrope Launched

Scope & strategy defined

2013

Initiate software development

Evaluation of existing standards

2014

Feasibility studies & POCs

Design, testing & due diligence

Phase I: Proof of Concept Studies

2015

API & Taxonomy development

V1.0 released internally (Sept)

1st deployments @ member companies

Phase II: Commercial Development

2016

ADF/API enhancements & testing

V1.1 released internally (Mar)

Increased vendor contribution

Release roadmap

V1.2 released internally (Nov)

2017

BioIT World Best Practice Award

Public Releases Q2, Q3, Q4

ADF Explorer/Viewer released

Business/Operating model 2.0

Vendor implementations

Full time resources & funding within vendors

Allotrope on vendor product roadmaps

Embedded in member companies; in production @ 2

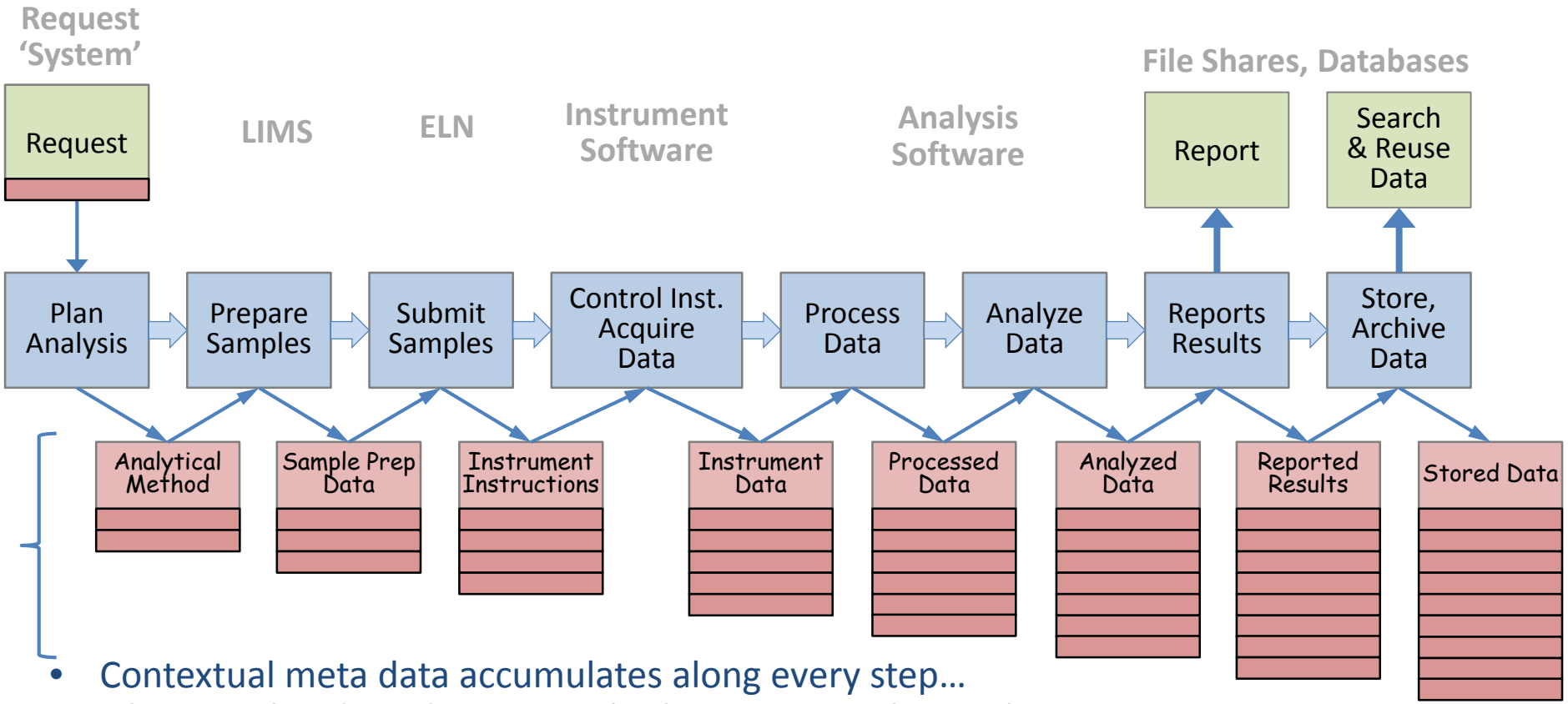
2022+

Phase III: Grow & Sustain Allotrope Ecosystem

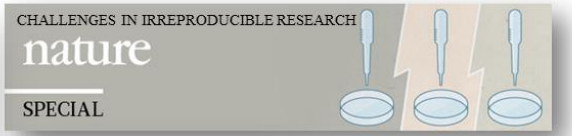
The Allotrope Community Today

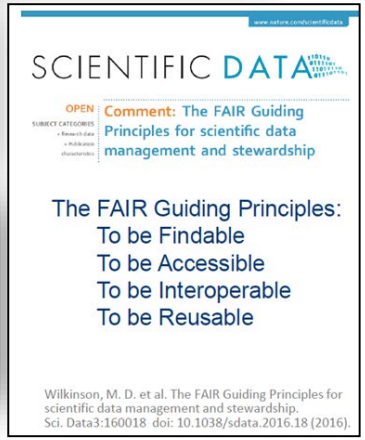
Astrix Technology Group • BSSN Software • Elemental Machines • Erasmus MC • Fraunhofer IPA • The HDF Group • LabAnswer • LabWare • Mettler Toledo • NIST • SciBite • Stanford University • University of Illinois at Chicago • University of Southampton

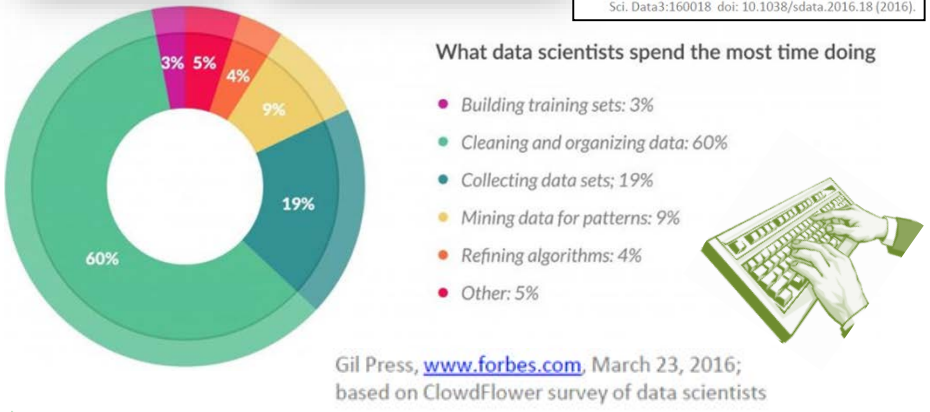
Fundamentals of a measurement workflow



- Contextual meta data accumulates along every step…
- …but it is distributed across multiple systems and records







Status quo in the laboratory

Data capture, integration & sharing challenges

- Some records still paper-based
- Manual transcription of methods and data
- Incompatible instruments & software
- Data integrity and scientific reproducibility challenges
- Legacy architectures are brittle & rigid
- Knowledge & context only in people's heads
- Silos of data, context and meaning
- Suboptimal knowledge management

Potential to delay getting medicines to patients & erosion of public confidence

From Eric Little, Osthus

Rethinking Scientific Data

What we do here…

…impacts everything here

What science needs…

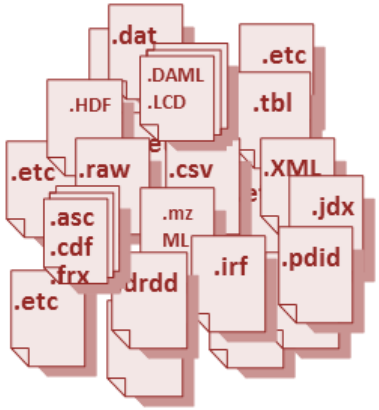
- A consistent way to record what we **observe** so we can
- **Share** findings with others so they can review, leverage, or
- **Repeat** our work

…what becomes possible

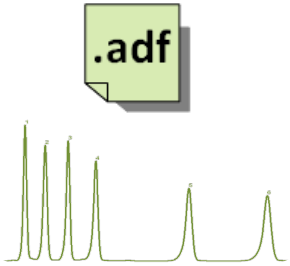
- Find any data in seconds
- Be 100% confident in the data integrity, quality and compliance of your results before you make decisions
- Analytics solutions using pristine metadata
- Share your science in a machine-executable format

Addressing the root causes

Vendor/instrument-Specific Formats



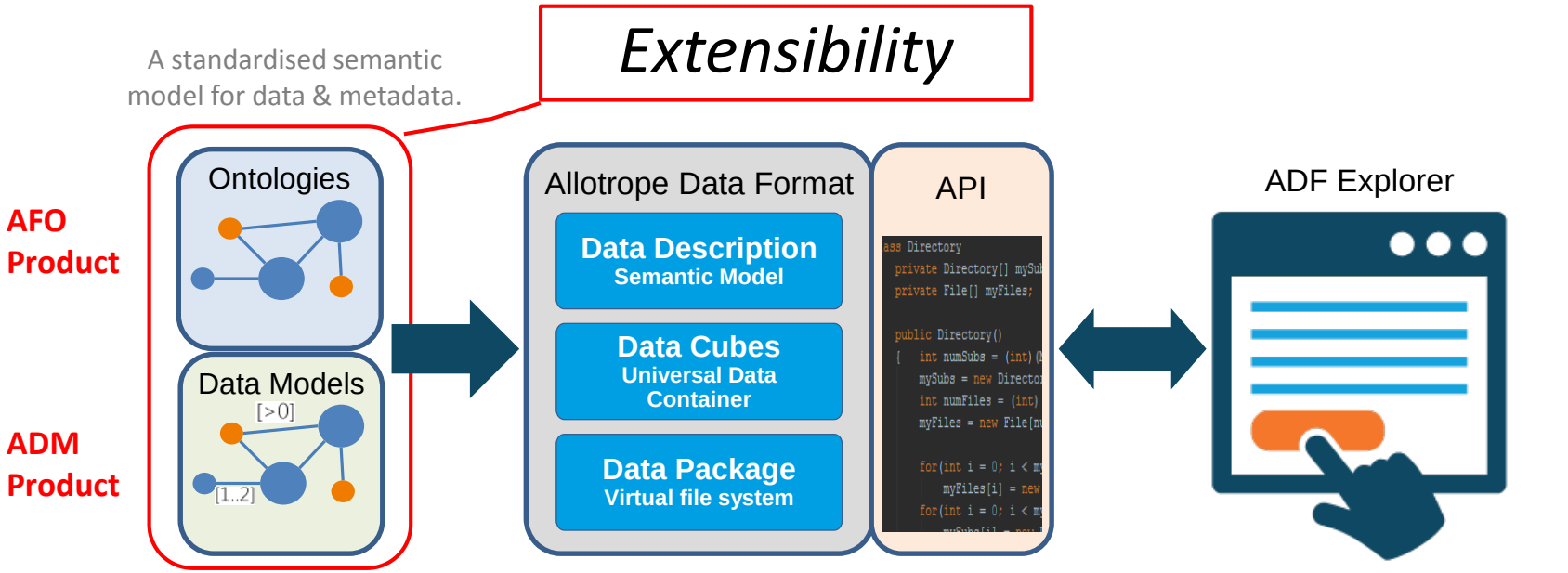
Data in Standard Format



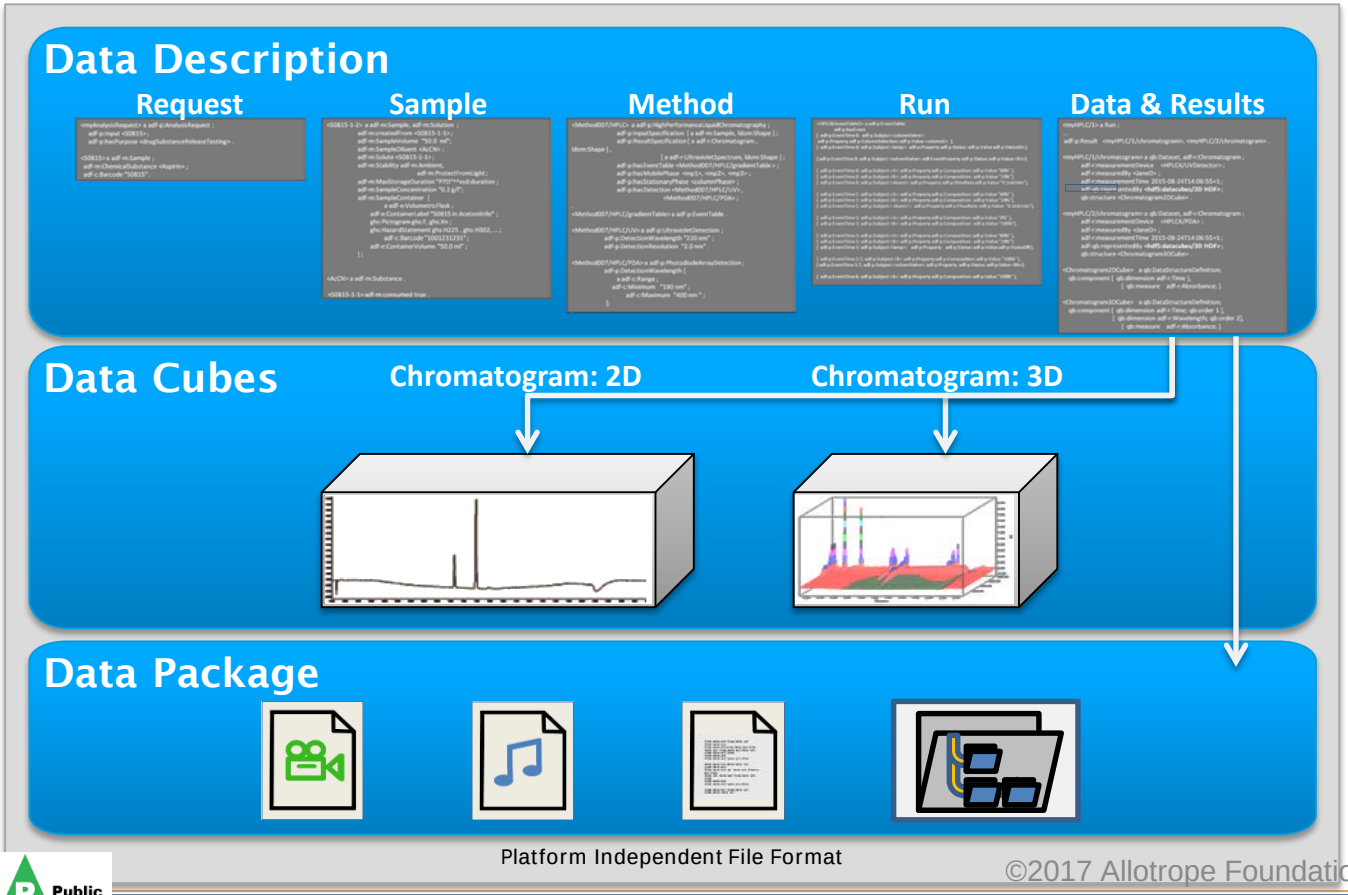
Paper-based and unstructured text for methods, regulations, recipes, observations, etc

A Standard vocabulary & metadata structure

The Allotrope Framework: Three Products, One Holistic Solution



Allotrope Data Format Example



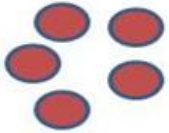
Descriptive metadata about
- Method, instrument, sample, process, result, etc.
- Data Cube, Data Package contents
- Provenance, audit trail, data models

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.

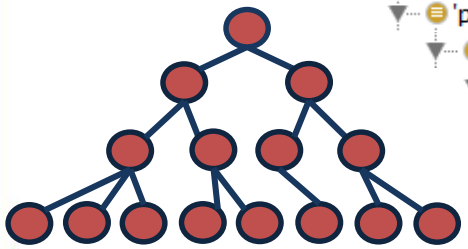
AFO Product Suite

Terminology



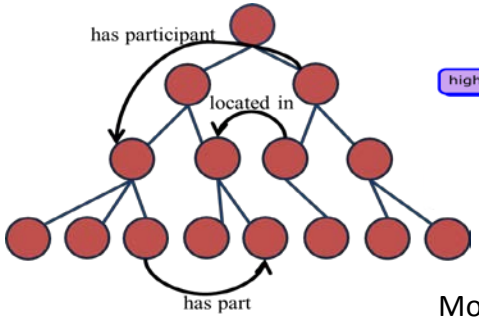
chromatography
column chromatography
liquid chromatography
chromatogram
peak
spectrum
ultraviolet spectrum

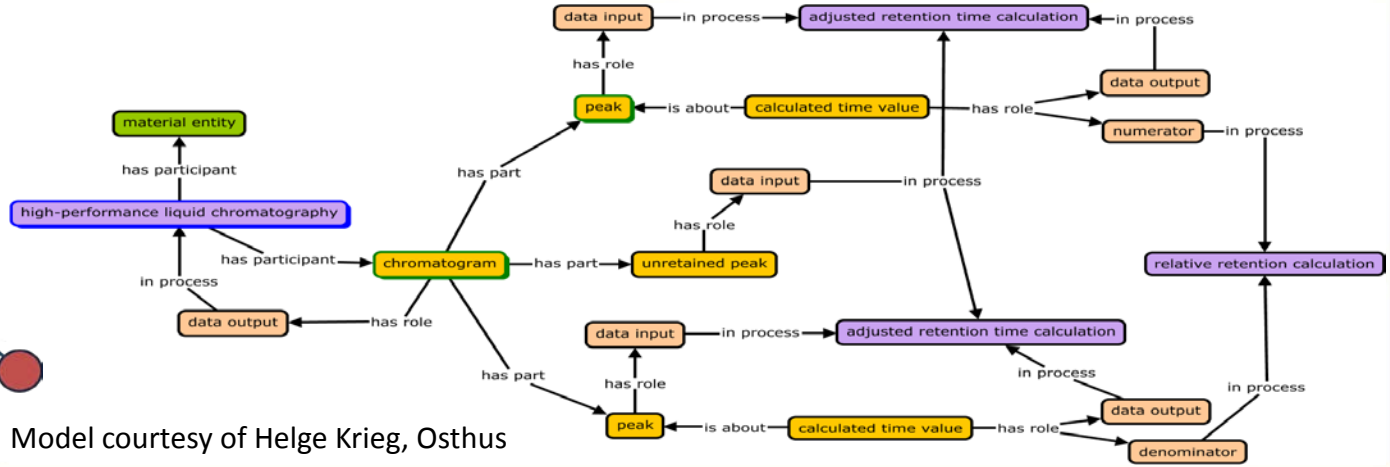
Taxonomy



- occurrent ≡ 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

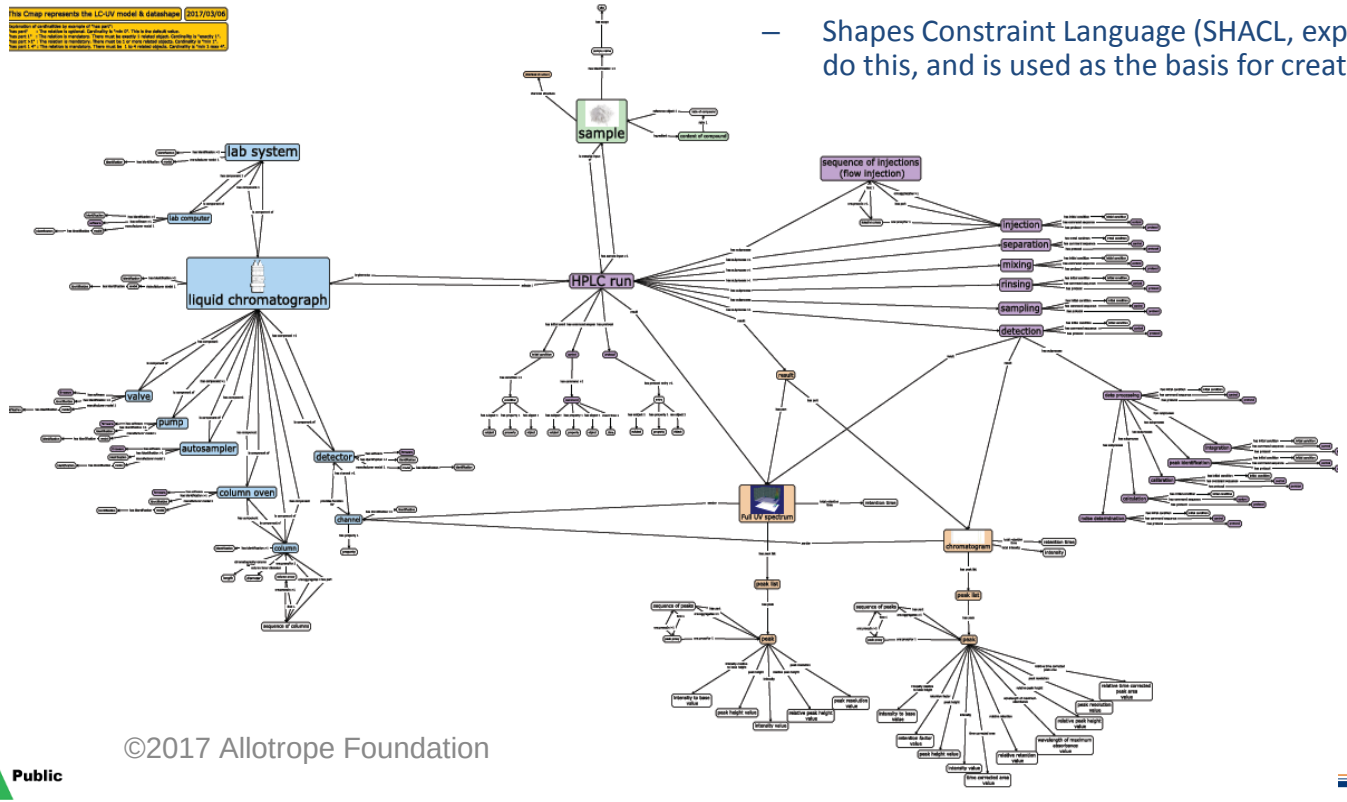




Model courtesy of Helge Krieg, Osthus

Allotrope Data Models: Standardized Use of Ontology in ADF Files

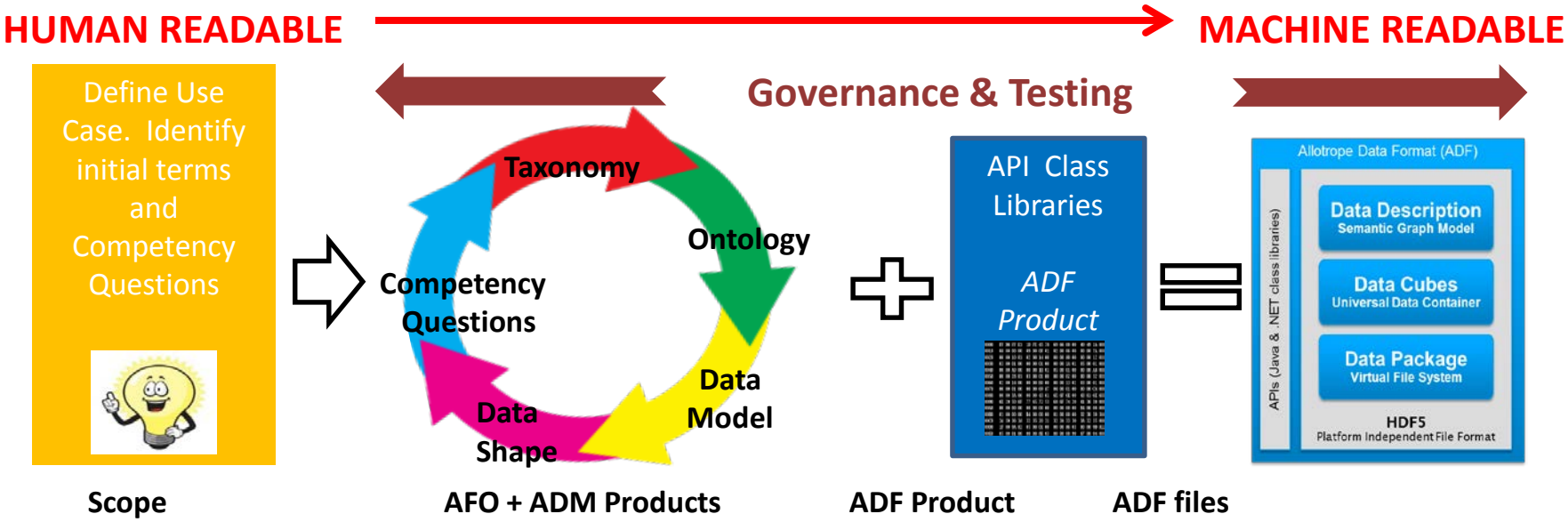
- Ontologies provide a consistent vocabulary we can use to describe things (instances) in our open world and give them a meaning (= **what** it is)
- Data structures (schemas, templates) describe **how** to use the ontologies for a given purpose in a standardized (reproducible, predictable, verifiable)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



Example: Portion of a CMAP graphical representation of LC-UV data model which shows the data structure using Allotrope ontologies.

The Allotrope intent is to partner with the Allotrope vendor community to create a consensus ontology and data model for use with all LC-UVs.

Reproducible Workflows to Create Ontologies and Data Models



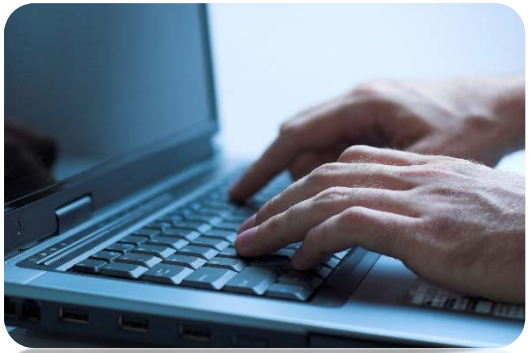
Competency questions link the laboratory scientist to the semantic engineer and software developer to express what information we wish to extract from our data

Anticipated benefits of applying the Allotrope Framework

Reduced Manual Effort & Paper

Better Scientific Reproducibility

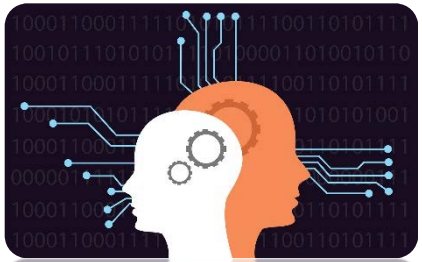
Increased Data Integrity, Context, Quality



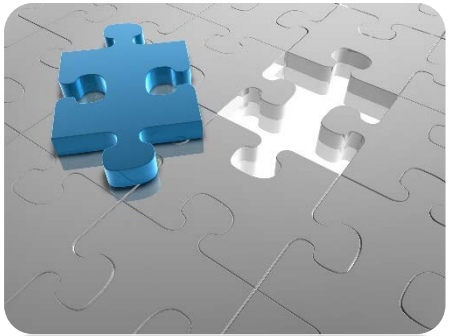
Simplified IT

Streamlined Access, Sharing, Integration

Foundation for Data Science



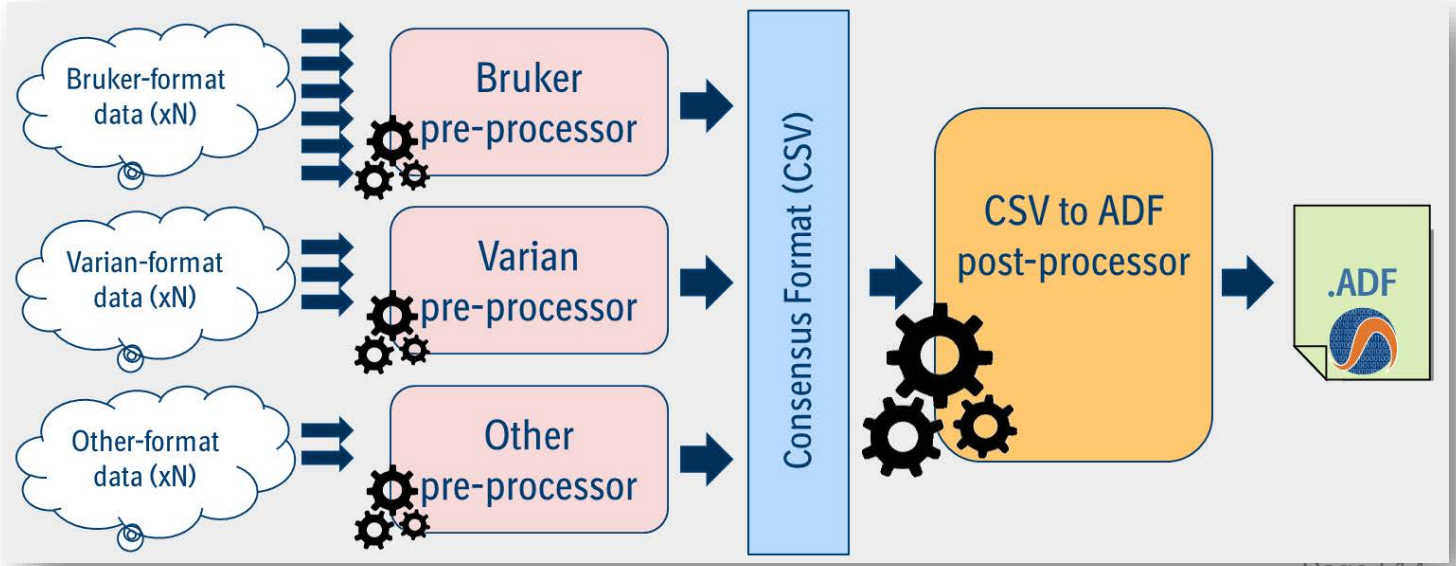
Consolidated Requirements Lower Innovation Barrier



Example Application. Simplified NMR Data Archiving

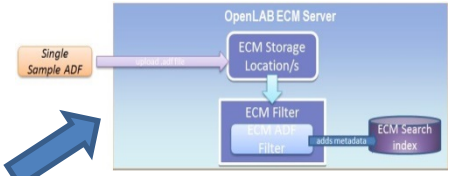
Boehringer Ingelheim

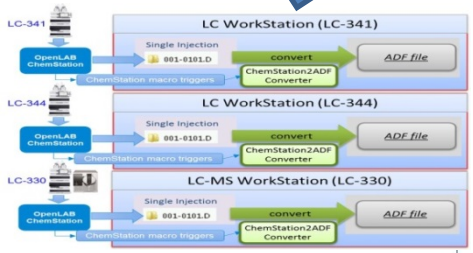
- Extensible architecture for vendor agnostic support for NMR
- Workflow and the applications to automate conversion of existing NMR data to ADF
- Successful population of ADF Data Cubes and the Data Package with content from legacy NMR data.
- Development of a “refresh” concept to populate the Data Description once semantic components of sufficient maturity are available.
- Demonstrated bi-directional, byte-for-byte, data conversion to/from ADF with zero information loss.



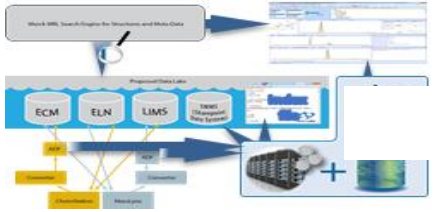
Example Application. Chiral Methods Screening E2E Workflows

Partnership with Agilent to create ADF filter pack to extract key meta data for indexing in ECM at archival

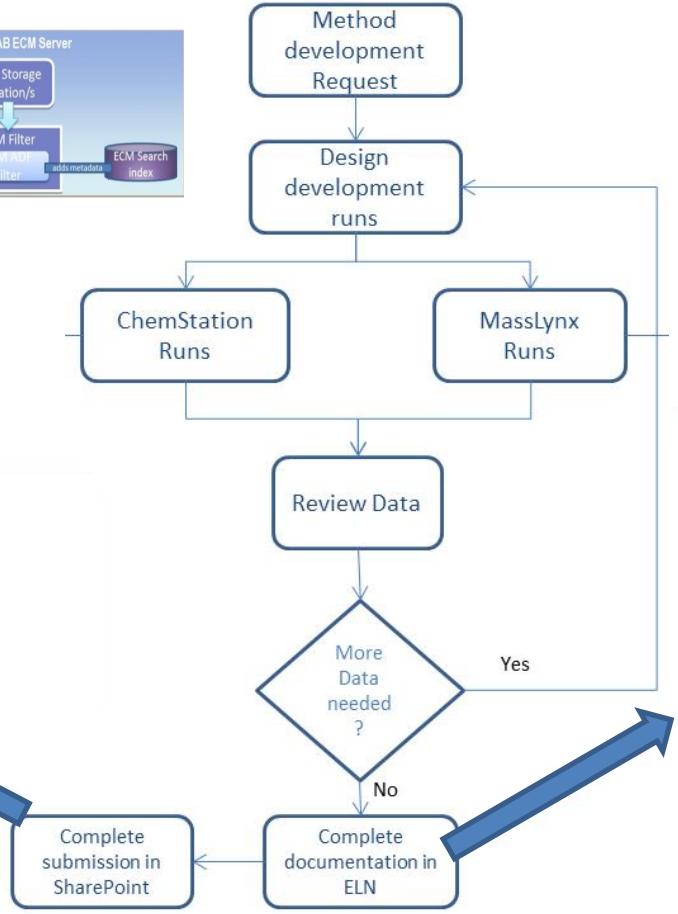


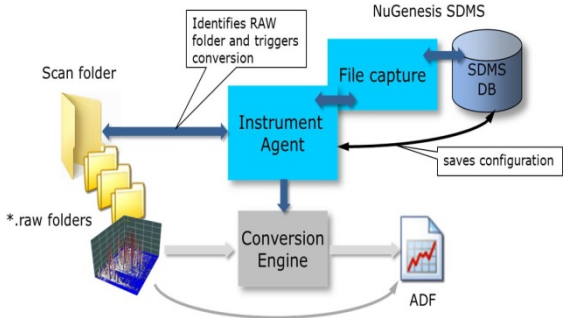


Partnership with Agilent to create ADF files using LC-UV ontology and data model

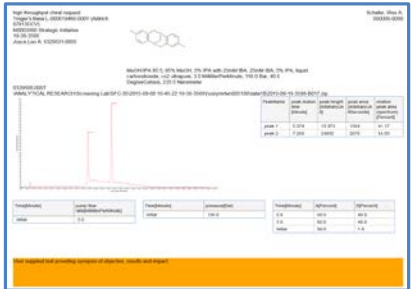


Partnership with ACD Labs to create automation server to add meta data to existing ADF files





Partnership with Waters to watch directories and create ADF files with Nugenesis SDMS using LC-UV ontology and data model



Partnership with Osthus (using Pipeline Pilot) to create ELN report

Example Application. Enterprise-wide data organization

Genentech
A Member of the Roche Group

Key Deliverables of Ana, Our Organization Platform

Standardization
Consistent Formats for Consistent Handling

Flexibility
Capable of Capturing Varied Data Types

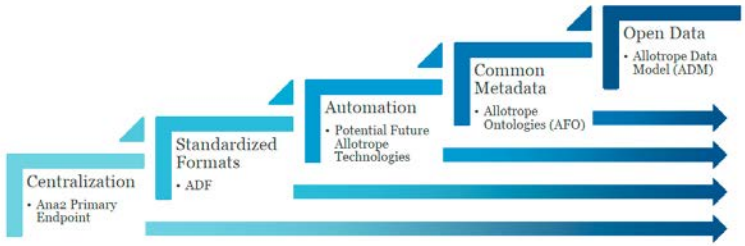
Scalability
Capable of Simultaneous Performance on 100+ instruments

Evolution
Able to Grow with the Data as it Evolves

- Single, Centralized Data Resource (+CRO data)
- Potential for Cross-platform Integration and Data Mining
- Enables Automation,
- Decreases Costs
- Increases Throughput

Genentech 8

A Layered, Evolutionary Approach to Data Organization



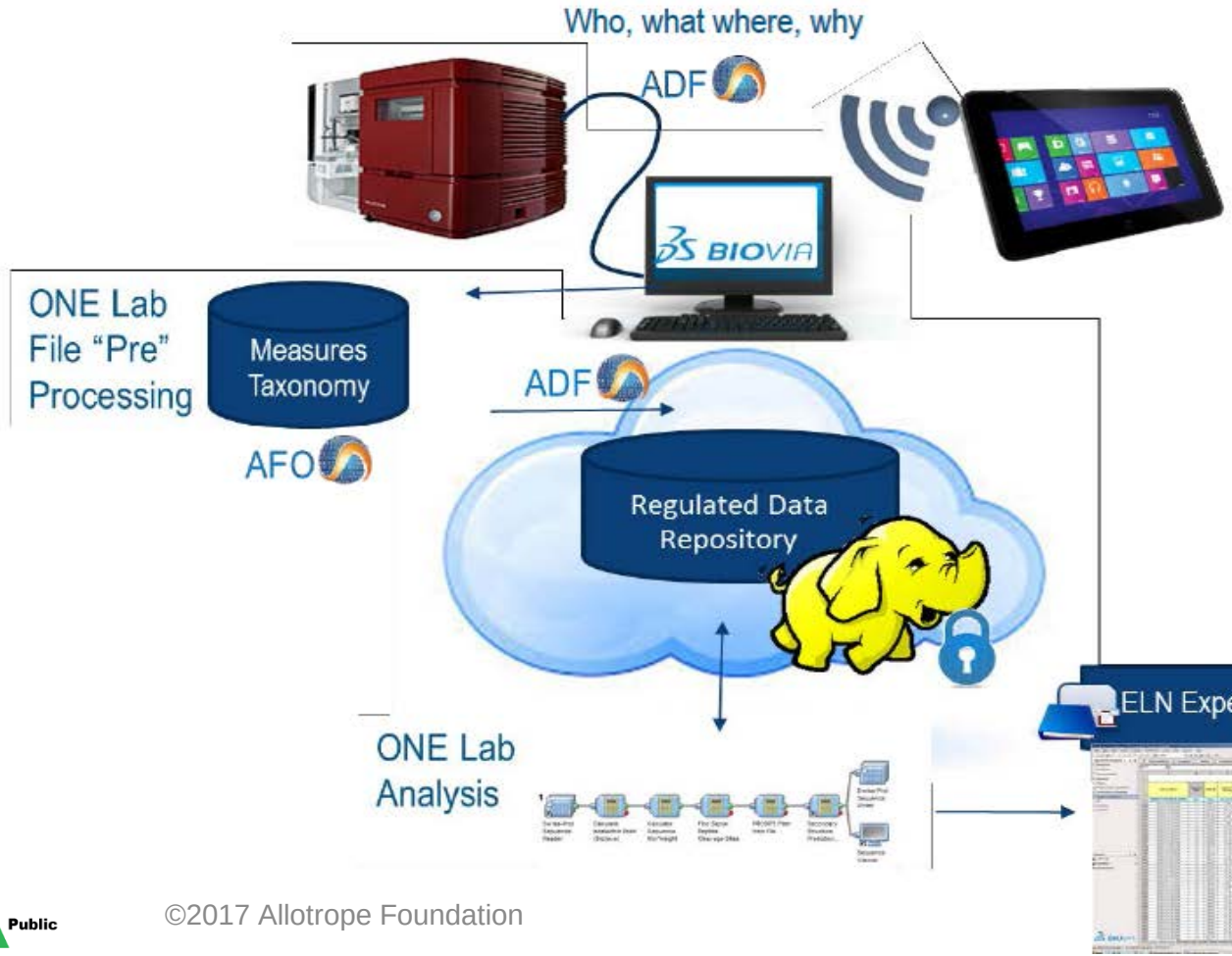
Our Approach: Deploy ADF Widely to Corral Data and Layer on Open Data as ADM Evolves

Genentech 11

Allotrope Framework provide a standard approach to…

- format analytical data during data capture
- label analytical data with consistent metadata during data capture
- link well-index analytical data to information in other repositories, such as substance registries
- But what else can we do with the Allotrope framework in the future ?

Example Application. Creating a Holistic Laboratory



Allotrope Foundation

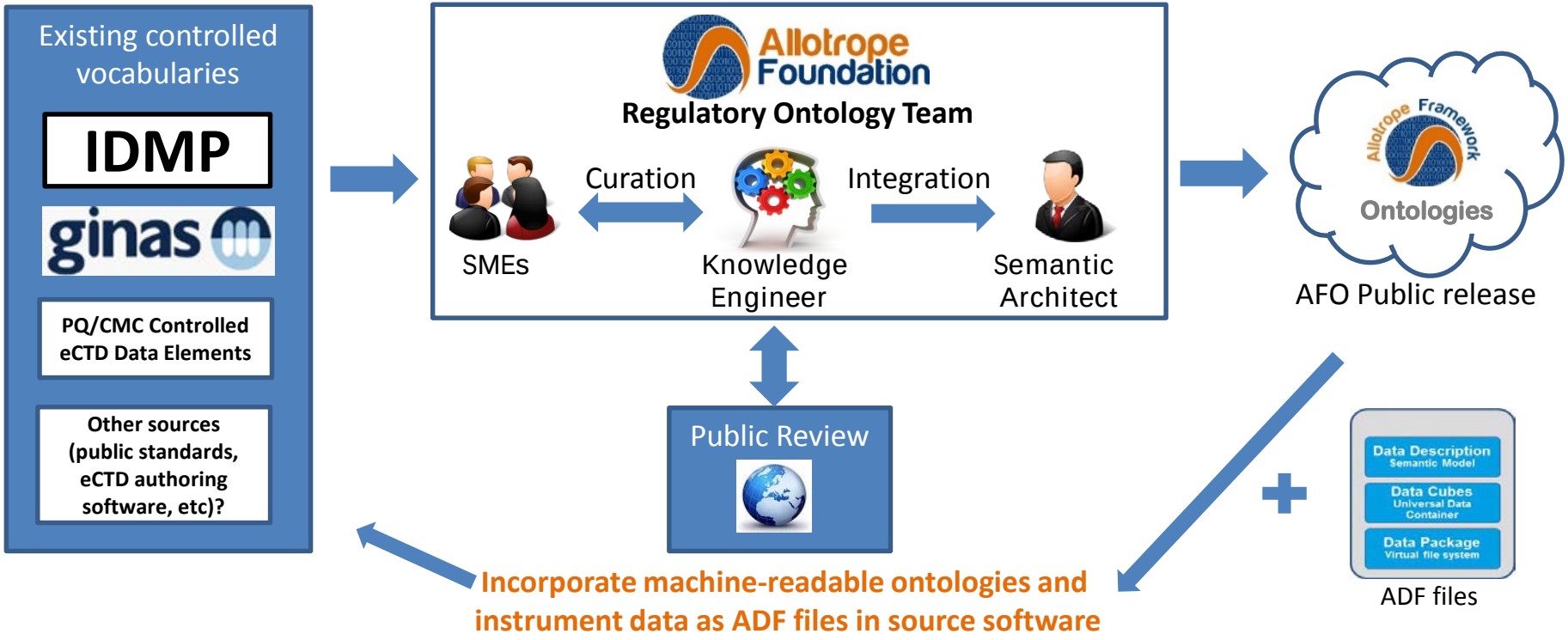
- BIOVIA foundational management of Allotrope Ontologies & Vocabularies
- BIOVIA foundational connection to laboratory instruments using Allotrope ADF format
- BIOVIA Recipe and Methods in S88 format
- Customer focused proof of concept projects in process and in the works

BIOVIA

AMGEN®

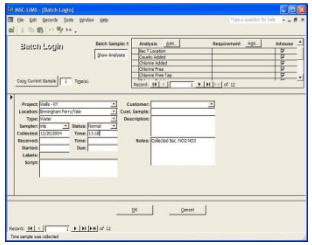
Public

Future Opportunity: Creating a Formal, Publically-Available Regulatory Ontology



Allotrope Regulatory Ontology Team composed of industry SMEs, regulators, and members of standards bodies

Future Opportunity: Creating an eCTD in Allotrope Format



Allotrope-compliant source information

automated transfer with 100% data integrity

eCTD data inputs standardized…

- ✓ ontologies
- ✓ Instrument data
- ✓ methods

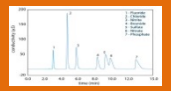
Allotrope-compliant eCTD authoring software

Data Description
Semantic Model

Data Cubes
Universal Data Container

Data Package
Virtual file system

Format eCTD in ADF container with associated data; automate data transfer where possible



leverage ontology & data integrity

eCTD data re-use standardized…

- ✓ ontologies
- ✓ Instrument data
- ✓ methods

reduced review & audit

better data mining

better trending & analytics

ginas

better downstream data interoperability

Back up slides

Join Allotrope Foundation!

Influence direction of development

Be part of an expanding community of experts

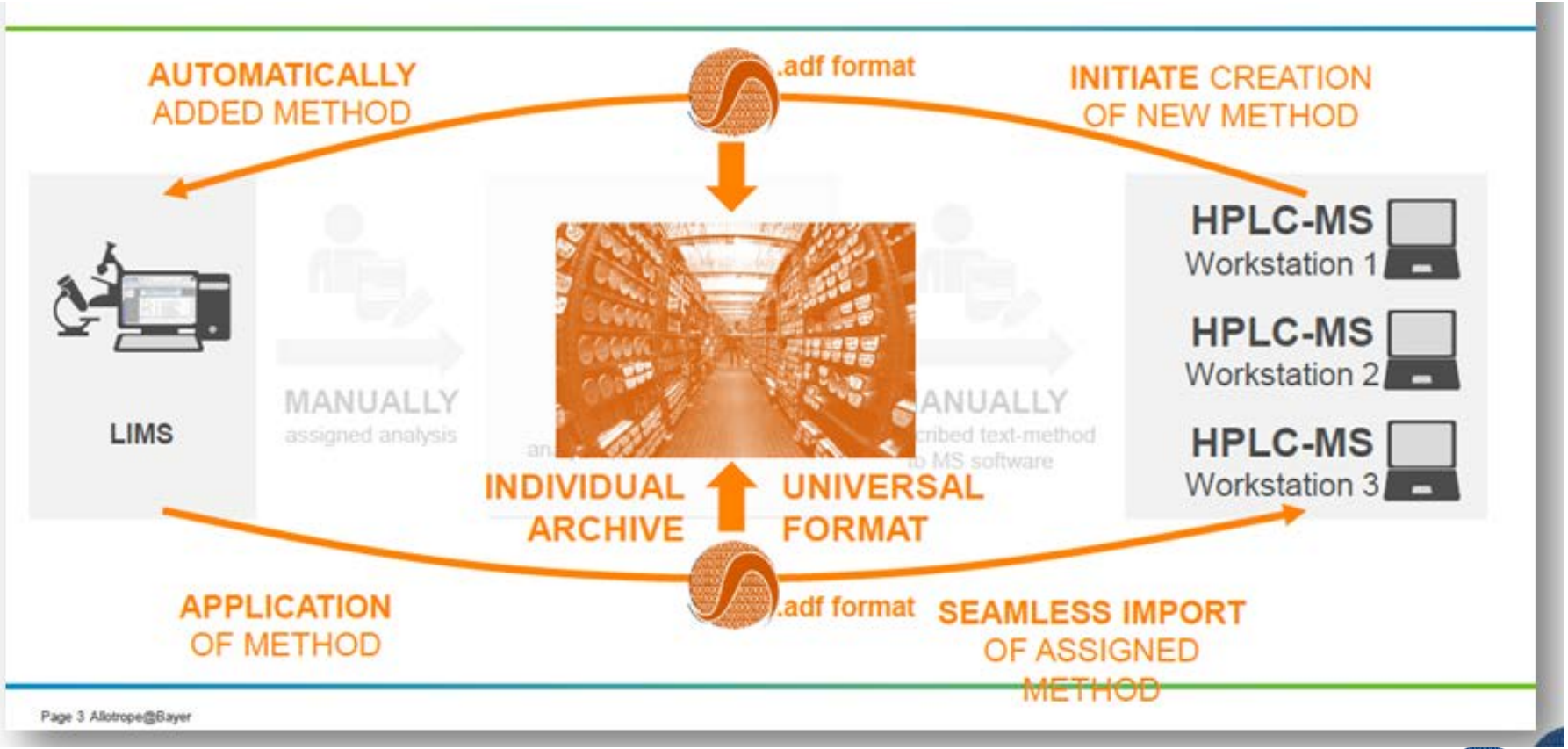
Receive support & training

Benefit from shared Investment

Align internal strategy with the future of data

Ensure sustainability & adoption

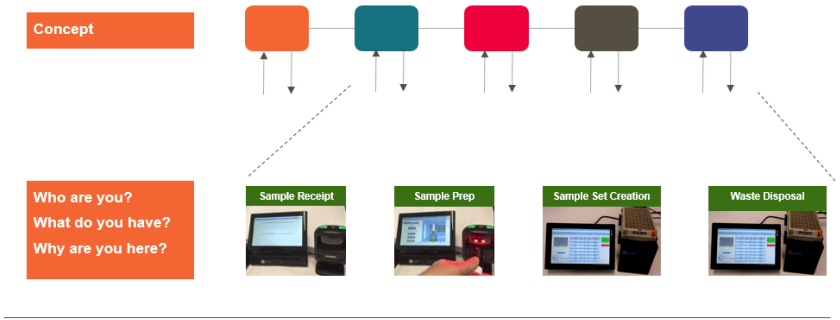
Example Application. Analytical Methods Management



Example Application. Keyboardless Linked Laboratory

Modularity is Key

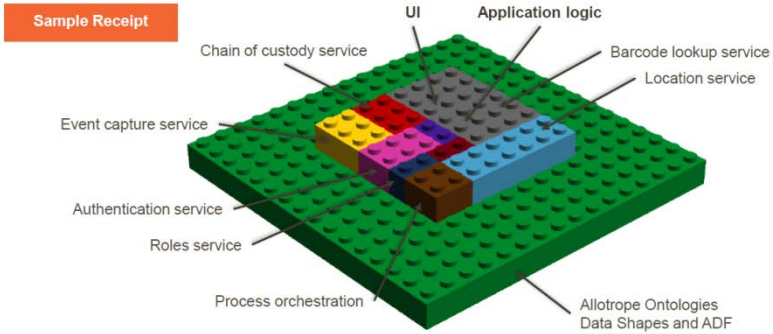
Data Stations use the Allotrope Framework to drive daisy-chained Unit Operations



Allotrope Framework > GSK Integration Project 12

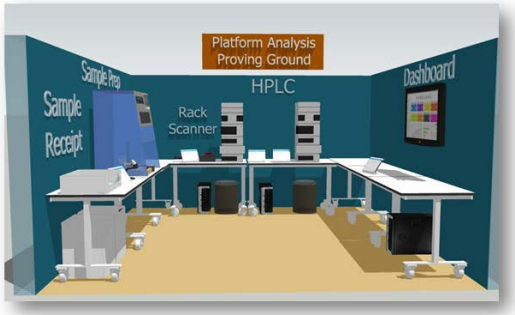
Modularity is Key

Modular means re-usable

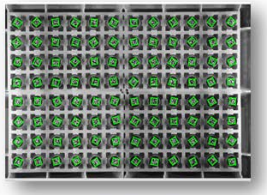


We created a lab to test the Framework

a place to seed and grow the vision



materials and data "flow" seamlessly from receipt to data preservation

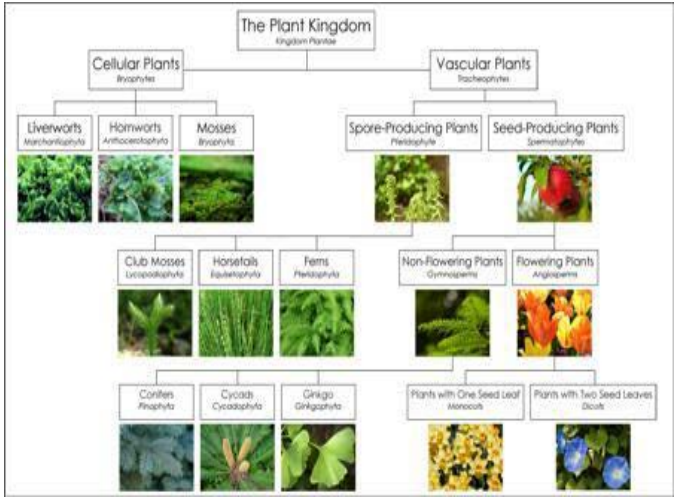


Ziath / Biosero Plate Reader

Relationship of Taxonomies and Formal Ontologies

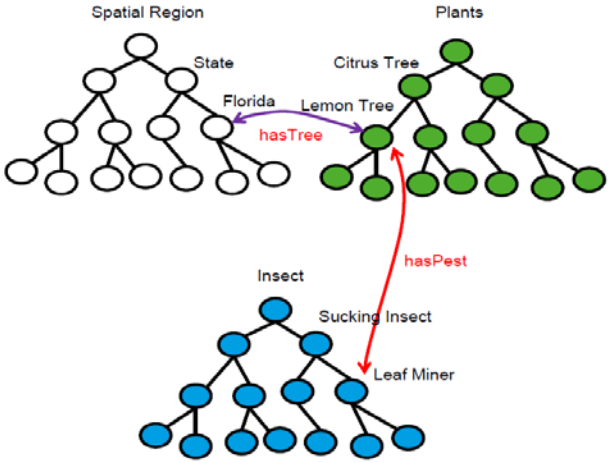
Taxonomies

Represent data as tree graphs and express fairly simple relationships (parent – child). It is critical to establish these relationships and definitions correctly as the basis of an ontology. **Taxonomies typically used in drop-down menus.**



Formal Ontologies

Captures human knowledge in a machine-readable way to create deeper, unambiguous relationships between data elements (groups taxonomies). **Formal ontologies facilitate use of computer reasoning and inferencing to perform more complex data queries and analytics.**



Allotrope Formal Ontology Alignment with the OBO Ontology Ecosystem (Basic Formal Ontology)

Allotrope Foundation Ontologies (AFO) Suite