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
Simplified IT

Better Scientific Reproducibility

Increase Data Integrity, Context, Quality

Reduce Manual Effort & Paper

Streamlined Access, Sharing, Integration

Consolidate Requirements Lower Innovation Barrier

Foundation for Data Science

©2018 Allotrope Foundation
Standardized Data & Metadata

How do we deliver this?
The Allotrope Framework

A standardised semantic model for data & metadata.

Ontologies

Data Models

Released Dec 2018

Released Jan 2019

Released Dec 2017

Released July 2017

Allotrope Data Format

API

Dec 2018

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

Data Description
Semantic Model

Data Cubes
Universal Data Container

Data Package
Virtual file system

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

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

©2018 Allotrope Foundation
Allotrope Data Format Example

Data Description
- Request
- Sample
- Method
- Run
- Data & Results

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

Data Cubes
- Chromatogram: 2D
- Chromatogram: 3D

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

Data Package

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

**HPLC-run**  
- Technique: HPLC
- Executed by user
- LIMS ID
  - Meas. Request IDs, not part of AFT yet
- Methods
  - Acq. Name, Last Modified DateTime
  - DA: Name, Last Modified DateTime
- Acq Method Parameters
  - converted a significant amount of acquisition parameters into the "<of-p>" "process" section using ATF

**Sample**  
- Sample name
- Sample description

**HPLC Instrument**  
- Instrument identifier: <LC 0312>
- Instrument description: <...>

**Chromatograms**  
- Identifier: <DAD1A>
- Created DateTime: <2015-09-18T23:02:44Z>
- Peaks ()
  - Index: <1>
    - Compound Name: <Trazadone>
    - Peak Elution Time: <1.20...> + Unit
    - Peak Area Value: <1.36...> + Unit
    - Peak Area Percent: <0.40...> + Unit
    - Peak Height Value: <21.22...> + Unit
    - Peak Height Percent: <3.06...> + Unit
    - Peak Start: <1.18...> + Unit
    - Peak Stop: <1.22...> + Unit
    - Peak Width: <0.02...> + Unit
    - Baseline value at Start of Peak: <1.10...> + Unit
    - Baseline value at End of Peak: <1.10...> + Unit

Links to data in ADF data cube
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
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?

**How?**

- CDS acquisition method UV-2345

**Why?**

- stability study

**What?**

- S-001-03-018

**Who?**

- SUB481
- ANA012

**Where?**

- site-1/building-B10/laboratory-E14

**When?**

- 2018-09-26T15:29:54.007Z

**What?**

- e3c388123-d091-dff3-b545-37260e8d1a82

**Related Patterns**
The Allotrope Ecosystem

Public Repositories?
E.g. http://obi-ontology.org

Allotrope Repository
https://allotrope.jfrog.io

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

Automated QC for Ontologies
Automation Infrastructure

Data Viewers
Data Validators
Data Model Editors
QC Tools

Open Source Tools

18 Data Models in Progress
Pistoia Method DB & Ontology Mapping

Collaboration Infrastructure

©2018 Allotrope Foundation
The Allotrope Community

Abbvie • AMGEN • Baxter • Boehringer Ingelheim
Bristol-Myers Squibb • Genentech • gsk • Lilly
MERCK • novo nordisk • Pfizer • Drinker Biddle

Allotrope Foundation

COGNITIVECHEM • The HDF Group • LABlicate
Malvern Panalytical • Tetrascience
Agilent Technologies • Mestrelab Research
Synthace

Allotrope Partner Network

Member

Cognizant • Cytobank • Waters

LEAP TECHNOLOGIES • PerkinElmer • Biovia

BIO-RAD

PANGAEA • SciEX

SHIMADZU • Waters

 Methodist University • NIST • SciBite • Stanford University
University of Strathclyde Glasgow • University of Southampton

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