

Driving Outcomes on The Foundation for the Digital Transformation of the Laboratory

Dana Vanderwall, Ph.D.

Director, Biology & Pre-Clinical Sciences IT; Bristol-Myers Squibb Chair, Allotrope Foundation Board of Directors

Rethinking Scientific Data

Contextual meta data accumulates along every step of a workflow...



Contextual meta data accumulates along every step of a workflow... ... distributed in pieces across multiple systems & records













Economies of scale?

- One study can encompasses multiple measurements
 - Diverse: different measurements & instruments
 - Longitudinal: long term studies/time points
 - Distributed: groups, departments or companies
- Technology has never been moving faster than it is today



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, <u>www.forbes.com</u>, March 23, 2016; based on ClowdFlower survey of data scientists



Allotrope Foundation, BioIT World 2017





Allotrope Framework

Standard Vocabulary



A holistic approach in a modular & extensible architecture, leveraging existing standards



Holistic framework to produce standardized & semantically aware scientific data









Allotrope Foundation

BFO Basic Formal Ontology CFR21 FDA Code of Federal Regulations Title 21 ChEBI Chemical Entities of Biological Interest CHEMINF Chemical Information Ontology CHMO Chemical Methods Ontology CL Cell Ontology CMO OBO Clinical Measurement Ontology DOLCE Descriptive Ontology for Linguistic and Cognitive Engineering

FDA US Food and Drug
AdministrationInve
PATIAO OBO Information Artifact
OntologyOnto
Onto
Onto
IUPAC International Union of
Pure and Applied ChemistryPSI-
Initia
RO (
VIVO
Sem
Wik
EncyMSI Metabolomics Standards
InitiativeWik
EncymzML Mass Spectrometry
XML format StandardWik
EncyNCI National Cancer Institute
Standards and TechnologyWik
Ency

nmrML Markup Language for NMR

OBI Ontology for Biomedical

Investigations **PATO** OBO Phenotypic Quality Ontology

PSI-MS Proteomics Standards Initiative

RO OBO Relations Ontology VIVO-ISF VIVO Integrated Semantic Framework Wikipedia Wikipedia, the Free Encyclopedia

Released at highest level of maturity or candidate

HPLC-UV, HPLC MS-SQD, gas chromatography, SFC-Chromotography, Raman, cell counting, blood gas analysis instruments, differential scanning calorimetry, pH meter, TGA, XPRD

Work in progress

LC-MS /Quantitative MS, conductivity, osmolality, pH, BGA, materials, NMR, LC-UV method model, particle size, dynamic vapor sorption

+60 additional domain modules at lower maturity levels sufficient for basic data annotation

The Allotrope Framework

A holistic approach in a modular extensible, architecture, leveraging existing standards





The Integration of Results, Data, & Meaningful Context



Implementing the standards



















Interoperable data for seamless sharing & integration; unambiguous conditions & parameters in a standard vocabulary and vendor agnostic format









Data is reusable & interoperable; future proof from obsolescence.

Why just search when you can find? The annotation is complete & consistent across instrument, lab, group, department, or company?



Any of these can be a starting point for your transformation...



Sounds

complicated...

Less so than the current

state

Standardize at the source, simplify everything downstream

Consider the alternatives

1) Status quo 2) Other Standards?



Modular architecture-- separation of format, controlled vocabulary, models for extensibility w/o breaking changes

API's provided as part of deliverables; no ambiguity in

Vocabulary useful in variety of applications & solutions, enabling higher quality metadata, construction of knowledge graphs, search indexing

Vocabulary that integrates with other references & knowledge sources in life science and laboratory ecosystem

Semantic representation of metadata enables use of off-theshelf technology to query, mine and parse metadata, leverages existing stds & best practices



Binary data format optimized for speed and efficiency

Integration of raw data storage with semantic capabilities

Data models allow automatic enforcement and verification of conformance at time of creation and afterward: enable certification of software applications

Built on field-proven container e.g. HDF originally developed with NASA, in use for ~25Y

Designed to support complex data formats (MS/MS, 2D NMR); general design for requirements of future data types

Allotrope Other **Community Sustainability** Framework standards Х Х experience, tools.

Based on real world use cases, input drawn from large community of analytical SMEs that know how instruments & data are used, as well as technical experts from instrument manufacturers & software companies

Not one person, or one company-backed with resources and funding by 14 industry leading pharma/biopharma companies and >50 partners

Support, maintenance and governance models established, with staff and funding model

Strategic direction and roadmap established in collaboration between pharma/biopharma leaders and partner ecosystem

Large, growing, active community; sharing expertise,



You'll get used to thinking about vour data in a different way.

APN Forums: Allotrope Members to privately meet with individual APN member companies

For non-confidential discussions around such topics as vendor product roadmaps, pharma data strategy, technology needs, requirements gathering, etc.

Wednesday, October 9

Room 42-1E

- 11:30 11:55 **Synthase**
- 12:00 12:25 Agilent
- 1:30 1:55 **Shimadzu**
- 2:00 2:25 **BioRad**
- 2:30 2:55 **TetraScience**

Thursday, October 10

Room 42-1D

- 2:00 2:15 Agilent
- 2:20 2:35 **Shimadzu**
- 2:40 2:55 **BioRad**

Room 42-1E

- 2:00 2:15 **TetraScience**
- 2:20 2:35 **Synthase**

