

Allotrope Data Format – Semantic Data Management in Life Sciences

Heiner Oberkampfl¹, Helge Kriegl¹, Christian Senger¹, Thomas Weber¹ and Wolfgang Colsman²

¹OSTHUS GmbH, Eisenbahnweg 9-11, 52068 Aachen, Germany

²OSTHUS Inc., 1990 West New Haven Ave, Melbourne, FL 32904 Florida, USA
heiner.oberkampfl@osthus.com

Abstract. Research intensive industries like the (bio-) pharmaceutical industry require a data architecture enabling universal laboratory data storage comprising describing metadata, ontologies and vocabularies, long-term data persistence, and metadata structure visualization. We present the Allotrope Data Format, the Allotrope Foundation Ontologies, ZONTAL Space and OSee as a comprehensive set of tools for laboratory data management.

Keywords: Allotrope, Ontologies, Data Model Patterns, Ontology Visualization

1 Allotrope Foundation and Allotrope Data Format

The **Allotrope Foundation** [1] is an international consortium of (bio-) pharmaceutical, and other research-intensive industries developing a data architecture for lab data acquisition, exchange, and management throughout its complete lifecycle. As the Allotrope Framework Architect, OSTHUS developed a unique framework for analytical data, consisting of a standard data format, class libraries for application interfacing, and semantic capabilities to support standardized, structured metadata. This framework has wide applicability across life science and other industries.

The **Allotrope Data Format** (ADF) is a universal data container based on HDF5 [2] with embedded semantics. It has three main components: (1) *Data Description*: Descriptive metadata for methods, instruments, sample, process, result, etc., as well as provenance information and audit trail for Data Description. (2) *Data Cube*: Analytical data represented by one- or multidimensional arrays of homogeneous data structures. A Data Cube preserves a vendor-neutral representation of the scientific data, without loss of information or precision. (3) *Data Package*: a virtual file system to store analytical data of arbitrary formats.

The **Allotrope Foundation Ontologies** (AFO) [3] provide the vocabulary for laboratory analytical processes. Its vocabulary is based on the Basic Formal Ontology (BFO) [4] reusing ontologies from the OBO Foundry [5]. AFO follows a modular approach of ontology development and covers domains of processes (e.g., HPLC), equipment, materials (extending ChEBI [6]), result specifications, roles and functions, and qualities.

To ensure interoperability, the **Allotrope Data Models** (ADM) [7] provide a set of semantic patterns, which are consistently reused across use-cases utilizing SHACL [8] for serialization. This allows for e.g., a comparable specification of experimental processes' results.

2 Long-Term Preservation of Data

Regulated industries require long-term preservation of data. ZONTAL Space [9] is the first archiving as a service platform that supports archiving of scientific and non-scientific data for the globally distributed enterprise. ZONTAL Space implements the Open Archival Information System (OAIS) [10] standard using ADF as the underlying data container for OAIS Information Packages (IPs). Utilizing the ADF semantic Data Description component, rich contextual metadata allows connecting data across organizational silos to create a holistic view on company's data assets and thus enables them to follow the FAIR Guiding Principles for scientific data management and stewardship [11].

3 Viewing Structured Metadata

Metadata stored in ADF is structured based on AFO vocabulary and ADM. A Key feature of ADF's structure of information is the interoperability of data being understandable to *machines*. OSee [12] is an open-source application that transforms machine-understandable data contained in ADF into a comprehensive, well-arranged graphical view in order to support understanding by human users.

References

1. Allotrope Foundation homepage <https://www.allotrope.org/>
2. HDF5 homepage <https://www.hdfgroup.org/solutions/hdf5/>
3. Allotrope Foundation Ontologies <http://purl.allotrope.org>
4. Arp, R. et al: Building Ontologies with Basic Formal Ontology. MIT Press, 2015
5. Smith, B., et al.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology* 25.11 (2007): 1251.
6. Degtyarenko, K. et al.: ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Research*. 2008
7. Allotrope Data Model Pattern Catalog <https://allotrope.gitlab.io/adm-patterns/>
8. Knublauch, H., et al: Shapes Constraint Language (SHACL), W3C Recommendation 2017, <https://www.w3.org/TR/2017/REC-shacl-20170720/>
9. ZONTAL Space homepage, <http://www.zontal.io/>
10. OAIS Reference Model (ISO14721), <http://www.oais.info/>
11. Wilkinson, M. D., et al.: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3 Article number: 160018 (2016).
12. OSee homepage <https://gitlab.com/allotrope-open-source/osee-server>