Allotrope Foundation Quarterly Update 2024/09

Dear Allotrope Community,

We have continued our progress this quarter and improved or expanded the AFO and ASM with updates to share. Please note that access to links may require access to GitLab or other Allotrope Community resources. More details for access here.

Table of Content

Allotrope Foundation Ontology & Data Models (AFO/ASM)

Modeling teams have continued working to align on proposals to expand the domain coverage of the AFO and ASM. Easily access files located on Client Connect here and more granular technical details available on GitLab, https://gitlab.com/allotrope.

See www.allotrope.org/product-releases for a full and updated list of available models.

AFO Updates

Following the new and updated set of ASM released this quarter, a new AFO release is published. Please note that QUDT 1.0 is no longer merged into the Allotrope Merged Ontology Suite.

The Allotrope Merged Ontology Suite release is available on:

- BioPortal, the repository of biomedical ontologies published by the National Center for Biomedical Ontology at Stanford University: https://bioportal.bioontology.org/ontologies/AFO
- OLS4, the Ontology Lookup Service repository for biomedical ontologies published by the European Bioinformatics Institute: https://www.ebi.ac.uk/ols4/ontologies/afo
- Ontobee, Ontologies data server published by the University of Michigan Medical School: https://ontobee.org/ontology/AFO (Ontobee generates the AFO list of terms in an Excel spreadsheet as well as Tab Separated Values file)
- Client Connect: here
- Gitlab: https://gitlab.com/allotrope/afo/-/tree/master/afo
- Allotrope PURL sever: http://purl.allotrope.org/ (listed under AFO>MERGED)
- JFrog Artifactory: https://allotrope.jfrog.io/ui/repos/tree/General/AFO-release-public
- Allotrope website: https://www.allotrope.org/ontologies

AFO Term Dictionary

Allotrope Term Dictionary is available in both .xlsx (Excel) and .csv (Comma Separated Values) format and can be downloaded from the

- Allotrope website at: https://www.allotrope.org/ontologies
- Client Connect: here
- JFrog Artifactory: https://allotrope allotrope.jfrog.io/ui/repos/tree/General/AFO-dictionary-release-public

Allotrope Foundation Simple Models (ASM)

ASM Directory

ASM Directory for the applicable sample JSON and JSON Schema files per technique is available for convenient viewing of file content using a browser. The directory contains links to the latest sample files and embedded (i.e., standalone) schema for all ASMs in REC (Recommended) status. Please refer to the directory on the public repository: https://gitlab.com/allotrope-public/asm/-/blob/main/README.md#allotrope-simple-model-directory

Modularization

JSON Schemas allow for modularization and factoring out commonly used rules by utilizing references to other JSON schema files. The simple model schemas make use of this modular approach. The ASM Schema is defined using:

- Technique specific schema: a JSON Schema that contains the domain specific rules. It references the core declarations instead of each technique defining its own.
- Core schema: a JSON Schema that contains reusable, domain independent rules. The core schema defines value types for all possible values that may be used in tabular models.
- Other reusable schemas: Cube, Hierarchy, Manifest, Units, other future extensions

Having the basic rules factored out in a core and other schemas allows for later extensions without changing each technique specific schema. It ensures consistent writing and querying regardless of whether it's a single contained instrument or a modular stack with multiple detectors, pumps, or anything else. Motivation of the modular pattern is to drive consistent data structures across techniques, enabling data from different models to work seamlessly together.

ASM Updates

New and updated Allotrope Simple Models release this quarter are available on:

- Client Connect: here
- GitLab: https://gitlab.com/allotrope/adm/-/tree/master/
- JFrog Artifactory: https://allotrope.jfrog.io/ui/repos/tree/General/ADM-release-public

The current Release Notes is available on Client Connect: here

Here is the list of the new and updated set of ASMs released this quarter.*

ASM Model	Type	Maturity	Path
Rheology	Tabular	REC	New
Light obscuration (as a Solution Analyzer detector)	Tabular	REC	Update
Automated reactors	Tabular	REC	Promoted from CR
Chromatography column	Tabular	REC	Promoted from CR
Electronic spectrometry (renamed from Spectrometry)	Tabular	REC	Update
GC-MS	Tabular	REC	New

* To find out how to access the related model's artifacts on GitLab: https://gitlab.com/allotrope/adm/-/wikis/Summary-Table-of-the-Governed-ADM-and-ASM-Techniques-Artifacts

ASM Training Materials and Working with the ASM

ASM training material is available on Allotrope public repository at the following locations:

- Brief introduction to ASM: https://www.allotrope.org/allotrope-simple-model
- ASM Primer: https://gitlab.com/allotrope-public/asm-primer/-/wikis/home
- ASM Jupyter Notebook Demo: https://gitlab.com/allotrope-public/asm-jupyter-demo It is a step-by-step example file for working with ASM files in a Jupyter Notebook. It was also tested with Google Colab.

ASM and ADM Modeling and Support

ASM related support tickets can be opened at the ADM project (https://gitlab.com/allotrope/adm/-/issues).

The Product Team can generate ADM specific artifacts (SHACL and its deployment using ADF) by request.

The latest updated set of ASM models is available on Gitlab. New and updated models will be released in conjunction with the release of new tabular models. Adopters can generate example results of tabularized data based on the JSON ASM format.

In cases where there is no tabular model for a chosen instrumentation type or technique, the product team is available to support the drafting of a new tabular model, and the Modeling Working Group is ready to review and govern drafted models.

Tooling, Testing, QA and Automation Pipeline

Further enhancements and automation of AFO QA

Further enhancements were done on the automated AFO QA tools using the CI pipelines:

- Refining regular expression for IRI checks. https://gitlab.com/allotrope-open-source/allotrope-devops/-/issues/253

Further enhancements to the ADM automated QA using the CI pipelines

Further enhancements were done on the automated ADM QA tools using the CI pipelines:

- Work in Progress: Check SHACL Shapes import existing AFO files. https://gitlab.com/allotrope-open-source/allotrope-devops/-/issues/251

AFO errors cleanups

- Work in Progerss: Semantic Error clean-ups to conform to Style-Guide: Vet new definition sources in Working Drafts (WD) https://gitlab.com/allotrope/afo/-/issues/637

- Updating QUDT entities without definition https://gitlab.com/allotrope/afo/-/issues/1029

General Maintenance

- Updates were done to the Ontobee annotations and auto ontology loading link from PURL
- Fixing release merged job handling of multiline literals
- Remove caches from the ADM and AFO GitLab CI pipelines
- Fixing metadata errors in AFO/QUDT
- Fixing reachability via cURL issues to some links on PURL server
- Detect failure to deploy to JFROG for AFO & ADM
- Aligning the AFO versioning with Git (maintain the current in-progress version in the develop branch at any given time). As a results, the AFO release process was updated and it is documented at https://gitlab.com/allotrope/afo/-/wikis/Releasing-the-AFO-&-ADM

We would like to thank Karin Colsman from the PharmaLex engineering team for her dedication and commitment to improve the overall tooling, testing, QA and automation pipeline.

Working Group Updates

Please note that the working groups meetings are recorded to improve access and transparency for those unable to attend or for the folks that are just interested in what's going on. To sign up for any working group, go to: www.allotrope.org/working-groups

Modeling: (Notes: here)

The ongoing modularization efforts within ASM continue across various working groups. The Common Hierarchy Schemas serve as a collection of "Lego-like" reusable building blocks, designed to ensure the creation of consistent hierarchical structures across different models. The Modeling Working Group was working on extending the Solution Analyzer model to support a Light Obscuration Detector, with new schemas developed: Light Obscuration Document and Distribution Document. Additionally, significant progress is being made in extending the Spectrophotometry Model (renamed Electronic Spectrometry) with multi-modal options, alongside the development of a Binding Affinity Model, which is an analytical technique used to assess and rank the strength of bimolecular interactions.

Chromatography: (Notes: here)

- The working group has officially released the recommended version of the Chromatography Column Model. This model extends the current chromatography column model integrated within the general LC framework.
- USP provided to AF for review a permission/license for its Chromatography Column Designation classes, also known as the LGS# (followed by a number indicating the specific Packing, Phase, Support needed to complete an analysis) within the AFO
    - USP Chromatography Database can be found at https://www.usp.org/resources/chromatographic-columns

Mass Spectrometry: (Notes: here)

We would like to express our gratitude to Graham McGibbon (ACD/Labs) for his continued leadership of the Mass Spectrometry (MS) Working Group.

This quarter, a new GC-MS model was developed and showcased by Lablicate. Special thanks to Matthias Mailänder from Lablicate for his dedication and contributions to this effort. Additionally, Lablicate has demonstrated ongoing progress with ADF/ASM plug-ins for OpenChrom utilizing MS data.

Plate Reader: (Notes: here)

The Benchling team is actively working to enhance the existing Plate Reader model to incorporate support for optical imaging and spectral scans. A draft model has been introduced to enable fluorescence detection, which includes capturing results with accompanying images.
Additionally, the Working Group's bi-weekly meeting schedule has been adjusted, and meetings will now take place on Thursdays at 11 AM EDT.

Flow Cytometry: (Notes: here)

The Flow Cytometry Working Group was initiated this year, and we are actively inviting members of the community to participate. The group is currently focused on defining the scope of the initial model, with an emphasis on reusing established standards such as FCS (Flow Cytometry Standard) and Gating-ML 2.0. Recent discussions include:

- Reviewing different gate shapes recorded by FlowJo.
- Capturing the compensation matrix within a Data Cube framework.

The team is now entering a stage where a tangible model for flow cytometry data can be constructed. A draft spreadsheet for mapping FCS Keywords to Allotrope Foundation Ontology (AFO) terms has been developed and is under review.

We encourage anyone who has publicly shareable flow cytometry data to contribute, as it would greatly assist in creating real-world examples to drive the initiative forward. Your input can help kickstart the next phase of our work.

Allotrope in the News

For the latest list of “Allotrope in the News”, please visit our website at:
https://www.allotrope.org/allotrope-in-the-news

Here is the listed recent news:

- Reimagining the Future of Biopharmaceutical Digitalization: Link1, Link2
- Be Open to Abstract Concepts such as Ontologies and Metadata Standards: Link1, Link2
- Xendat Data Deep Dive: Ontologies and Drug Discovery – Chemistry: Link
- MALDI-TOF MS Peaks: Link1, Link2

- OpenChrom Allotrope Plugins: Link1, Link2
- OpenChrom Allotrope Plugins explained: Link
- Data Centricity: The Key to Successful Digital Labs: Link
- Generating knowledge graphs through text mining of catalysis research related literature: Link1, Link2
- Introducing the 2024/Q2 Allotrope Release: New Solution NMR and Electrophoresis Models Plus More!: Link
- The landscape of ontologies in materials science and engineering: A survey and evaluation: Link
- A Controlled Vocabulary and Taxonomy for the Submission of Quality Attributes for Therapeutic Proteins Solutions: Link1, Link2
- Realizing the benefits of AI in Biopharma: Collaborative open source projects for data management: Link
- Agilent's Environmental, Social, and Governance (ESG) Report: Advancing Great Science with Sustainable Solutions: Link
- Digital Standards: A Path to Sustainable and Interoperable Chemical Data Exchange: Link
- SLAM – A thin-client for interoperable annotation and biomedical signal handling: Link
- Market Report: Biotechnology Instruments Market Size to Touch USD 132.80 Billion by 2033: Link1, Link2
- Allotrope plugin for OpenChrom has been published: Link
- Technology Trends in Separation Science: Data Handling: Link1, Link2
- NFDI4Cat 2.0: International Partners: Link

Projects within the Allotrope Community

Sample of Projects with the Allotrope Framework

- Lablicate released the Allotrope ASM/ADF plugin for OpenChrom. More information at the OpenChrome Marketplace. Read the blog post.

"The package includes read and write support for the Allotrope Data Format (ADF) and Allotrope Simple Model (ASM) so you can convert from and into any other OpenChrom supported file format. The ADF binary format also includes support for loading legacy raw data from the data package, so both can be used for processing in the Data Analysis main view. Additionally, an ADF explorer allows skimming through Allotrope Data Format files and discovering their ontologies in a table and graph view.

Supported techniques:

  - Allotrope Data File Format:
    - HPLC-UV/Vis and HPLC-DAD
    - GC-FID

        - GC-MSD (experimental)
        - MS (profile, read-only)
    - Allotrope Simple Model Format:
        - GC-FID (read only)
        - HPLC-UV/Vis
        - LC-MS (TIC only)
        - Realtime qPCR
        - FT-IR (read only)
        - Mass Spectrometry (single scan, read only)

- Shimadzu Supports the Allotrope Data Format (ADF) with LabSolutions:
    - LabSolutions is Shimadzu's software platform designed to manage data from Shimadzu's various analytical instruments, including chromatography, mass spectrometry, spectroscopy, and other laboratory equipment.
    - "The data obtained with LabSolutions can be output in a variety of formats. The system is even compatible with output in the data format specified by the Allotrope™ Foundation, an international consortium consisting of worldwide pharmaceutical companies, biotech companies, analytical instrument manufacturers, and software companies. Further, during data acquisition and data analysis, conversion to multiple data formats can be implemented automatically"
    - More information is available at:
        - https://www.shimadzu.com/an/products/software-informatics/labsolutions-series/system-linkage-option/index.html
        - https://www.shimadzu.com/an/products/software-informatics/labsolutions-series/system-linkage-option/features.html

- Merck Manufacturing Division (MMD) adopting ASM

Merck Manufacturing Division (MMD) started a pilot project utilizing the ASM as a standard across their instrument integration program

Instrument Data Converters to ASM (Open Source)

- Benchling is growing its platform for lab instrument connectivity and data management, Benchling Connect. With Connect, Benchling confronts industry-wide challenges with proliferation of proprietary instrument data models and vendor lock-in by mapping all instrument output to the Allotrope Simple Model (ASM) and making the converter codes open source and freely available on GitHub.
    - Benchling's blog: https://www.benchling.com/blog/benchling-connect
    - "allotropy", open source, Python library repository on GitHub: https://github.com/Benchling-Open-Source/allotropy

- "allotropy", Python package for converting instrument data into ASM is on PyPI: https://pypi.org/project/allotropy/
- Benchling Connect - Adapter specific guides: here
- Instruments supported by Benchling Connect: Detailed list by instrument category https://help.benchling.com/hc/en-us/articles/22558210727565-Instruments-supported-by-Benchling-Connect

Allotrope Publications

We have published 2-page summaries and updated the introductory presentation:

- **Allotrope Introductory Slide Deck**: can be downloaded from here
- **Allotrope Models & Domains**: can be downloaded from here
- **Allotrope Data Strategies**: can be downloaded from here

AF Community and Events

2024 Fall Allotrope Connect Workshop: Registration and Call for Presentation Abstracts

The next Allotrope Connect event, hosted by ThermoFisher, will be held on November 19th and 20th at the Marriott Long Wharf in Boston, MA.

- The registration is live: here
- 2024 Fall Allotrope Connect Workshop information page: here

Call for Presentation Abstracts: If you would like to present your work related to your use of the Allotrope Foundation technology, please complete this form by October 10th, or contact the product team.

Allotrope Data Framework Onboarding Guide

The Allotrope Onboarding Guide wiki page was updated. Please refer to the following link: Allotrope Data Framework Onboarding Guide

Allotrope YouTube Channel

Our YouTube channel has new a handle: https://www.youtube.com/@allotropefoundation. The Allotrope YouTube Channel hosts a technical playlist as well as the Allotrope Connect public presentations from 2017 and 2020 to the latest 2024 Spring Connect event.

The YouTube Channel videos are organized by playlists at:
https://www.youtube.com/@allotropefoundation/playlists.

Looking Forward

The Allotrope Product Team is looking forward to another productive 4th quarter of 2024. We are looking to develop additional improvements to meet the evolving needs of our community.

Please contact us for any questions at product_team@allotrope.org.

Sincerely,

Allotrope Product Team