Allotrope Foundation Quarterly Update 2024/12

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.

Allotrope Foundation Ontology & Data Models (AFO/ASM)

Modeling teams have continued working to align with 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
Plate reader (spectral scan)	Tabular	REC	New
Plate reader (optical imaging)	Tabular	REC	New
Electronic lab notebook	Tabular	CR	New
Electronic Spectrometry (spectral scan)	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

General Maintenance
- Aligning the AFO repository and the ADM/ASM versioning with Git (maintaining the current in-progress version in the development branch at any given time). As a results, the AFO and the ADM release process was updated, and it is documented at https://gitlab.com/allotrope/afo/-/wikis/Releasing-the-AFO-&-ADM
- Fixing reachability via cURL issues to some links on PURL server https://gitlab.com/allotrope-open-source/allotrope-devops/-/issues/24

Further enhancements and automation of AFO QA
Further enhancements were made to 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
- Updating QUDT entities without definition: https://gitlab.com/allotrope/afo/-/issues/1029
- Semantic Error clean-ups to conform to Style-Guide: Vet new definition sources in Working Drafts (WD) https://gitlab.com/allotrope/afo/-/issues/637

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

- Add format checking to ASM https://gitlab.com/allotrope-open-source/allotrope-devops/-/issues/232
- Format validation in the ASM JSON files is done using Python JSON schema validation tool https://python-jsonschema.readthedocs.io/en/latest/validate/#validating-formats
- Checking the ASM manifests import existing AFO files: New checks to the manifest validator, JSON scripts, to ensure the existence of the referenced .ttl files from the AFO. AFO entities checks was added https://gitlab.com/allotrope-open-source/allotrope-devops/-/issues/238 \
- Improving and automating ADM QA: Checking SHACL Shapes import of existing AFO files https://gitlab.com/allotrope-open-source/allotrope-devops/-/issues/251

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 modularization efforts within ASM are actively progressing across various working groups. The Common Hierarchy Schemas act as a collection of reusables, "Lego-like" building blocks, enabling the consistent development of hierarchical structures across different models.

Initiated by Benchling, Working Group was working on developing a Binding Affinity technique with Surface Plasmon Resonance (SPR) as the detection method. Binding affinity techniques are pivotal in measuring the interaction strength between molecules, such as drugs, and their biological targets, such as proteins. This process is essential in drug discovery and development, as it helps evaluate the efficacy of drug candidates in binding to their targets, optimize their potency, and reduce potential side effects.

Chromatography: (Notes: here)
The Chromatography Working Group focused this quarter on reviewing the modularization of chromatography models, including efforts to expand a common core for chromatography. The team also explored potential improvements in modularization inheritance to enhance the models' scalability and flexibility.

Looking ahead, the group anticipates work next year on several key areas, including Multi-Angle Light Scattering (Merck), Size Exclusion Chromatography (Merck), and Supercritical Fluid Chromatography (GSK). Additional efforts may involve incorporating gradients, such as:

- Temperature Gradient: Controlled increases in column temperature during the separation process.

- Solvent Gradient: Gradual changes in the mobile phase composition over time during separation.

These developments aim to further refine and standardize chromatography models.

Mass Spectrometry: (Notes: here)

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

This quarter, the team focused on adding ionization type terms to the ontology, addressing the lack of a well-structured ontological classification for ionization methods. Additionally, the team reviewed Lablicate's proposed draft for modeling MALDI-TOF (Matrix-Assisted Laser Desorption/Ionization Time of Flight), a mass spectrometry technique widely used for analyzing biomolecules such as proteins, peptides, DNA, and large organic molecules.

Plate Reader: (Notes: here)

The Benchling team has been actively enhancing the existing Plate Reader model to include support for optical imaging, enabling fluorescence detection with results presented as images, and spectral scans, extending absorbance and fluorescence models to cover scans across a spectrum of wavelengths. Looking ahead, the team plans to focus on incorporating support for area scans, a critical capability widely used in immunology and biology.

Flow Cytometry: (Notes: here)

The team has reached a significant milestone in developing a tangible model for flow cytometry data, utilizing established standards such as FCS (Flow Cytometry Standard) and Gating-ML 2.0. A draft spreadsheet mapping FCS Keywords to Allotrope Foundation Ontology (AFO) terms has been developed and reviewed.

The working group has been focused on resolving and aligning the ASM model by analyzing and applying example data from FlowJo and incorporating data regions to capture gating analysis.

Looking ahead, the first release of the flow cytometry model is expected next quarter, alongside future work on spectral cytometry (Merck).

ELN: (Notes: here)

In October, Allotrope launched an Electronic Lab Notebook (ELN) Working Group with the objective of standardizing the representation of ELN data using the ASM format. The group meets weekly on Fridays from 9:00 to 10:00 AM EST.

This initiative aims to expand domain coverage beyond its primary focus on the measurement domain into the experiment authoring domain. The first step involves developing a vendor-neutral data model to record ELN data in a consistent, human-readable ASM format. This model will serve as a general-purpose framework for capturing data in a standardized manner, including:

- Metadata: Recording essential information such as analyst names, timestamps, and locations using precise terms from the Allotrope Ontology for clear semantic definitions.

- Scientific Data: Capturing diverse scientific information through generic data patterns, further annotated with ontology terms to enhance searchability and usability.

This quarter, the team defined the scope of the initial model use case: Archiving complete ELN entries to a data lake. Merck and ZONTAL provided an initial model designed to capture this data. The model, while highly generic with minimal semantic details, served as the foundation for developing model patterns included in the current CR release.

If you or a colleague is interested in contributing to this initiative, or if you are a subject matter expert (SME), please feel free to reach out to the Allotrope Product Team for more details and involvement opportunities.

Projects within the Allotrope Community

Sample Projects with the Allotrope Framework

- **Merck Manufacturing Division (MMD) adopting ASM**

Merck Manufacturing Division (MMD) is working on a project utilizing the ASM as a standard across their instrument integration program

- **Bug Fix Contribution by ZONTAL to the h5ld Open Source Python Package**

h5ld is an HDF5 Linked Data project designed to facilitate user-created metadata within HDF5 files. This Python package provides readers for HDF5-based formats with linked data, enabling the entire linked data content to be loaded in one operation and made available as an rdflib graph object. Currently, the package supports the Allotrope Data Format (ADF). The project was originally developed by Aleksandar Jelenak from the HDF Group a few years ago (https://github.com/HDFGroup/h5ld).

ZONTAL contributed by identifying and resolving an IndexError bug within the package.

- **Scitara utilizes ASM with its DLX platform**
Scitara DLX platform supports three core functions: Connect, Automate, and Monitor in a single platform to support centralized management of all integrations and orchestrations across a diverse lab ecosystem. It Eliminate manual transcription errors, ensure data integrity, and standardize data collection and workflow practices for time and cost savings. Scitara can now implement FAIR Data Stream utilizing ASM.

- **Ganymede cloud-based data platform processes ASM**
Ganymede is a cloud-based data platform, engineered to streamline the capture and processing of data between lab instruments, ELN/LIMS/analytical applications. Ganymede can convert instrument data files into the Allotrope format using ASM mapper.

- **Lablicate OpenChrom plugins**

[Lablicate](#) is working on ASM plugin for MALDI-TOF (Matrix-Assisted Laser Desorption/Ionization Time of Flight), an MS technique used primarily for the analysis of biomolecules like proteins, peptides, DNA, and large organic molecules.

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. The Python project is called Allotropy. For an up-to-date list of available instrument converters please refer to the following GitHub page: https://github.com/Benchling-Open-Source/allotropy/blob/main/SUPPORTED_INSTRUMENT_SOFTWARE.adoc

---

Allotrope Publications and Media

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](#)

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.

Allotrope on LinkedIn

The Allotrope LinkedIn page is very active, with frequent updates and new posts. We encourage you to stay connected and follow us at: https://www.linkedin.com/company/allotrope-foundation

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 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:

- Allotrope's Role in Advancing FAIR Data with LADS OPC UA at NFDI4Chem: Link1, Link2
- Day 2 of the 2024 Allotrope Connect Workshop is Happening Now!: Link
- Robust "ELN in a future data-centric ecosystem" discussion now in progress at Allotrope Connect Workshop in Boston!: Link
- The 2024 Allotrope Connect Workshop is happening right now in Boston, hosted by Thermo Fisher Scientific: Link
- Streamlining Analytical Method Validation for ICH Compliance with ASM, eCTD Submissions, GAMP5 Qualification, and Standardized Data Integration: By PharmaLex at Fall 2024 Allotrope Connect in Boston: Link
- Instrument Integration to Knowledge Graph Development journey - To enable seamless flow of data in Laboratory Space with Allotrope: By Merck at Fall 2024 Allotrope Connect in Boston: Link
- How Agilent will use Allotrope Data Format to improve Laboratory Efficiency? - By Agilent Digital Lab Innovation Team at Fall 2024 Allotrope Connect in Boston: Link
- Enabling Automated End-to-End Chromatographic Data Workflows and Accelerated Data Insights with ASM Vendor-Neutral Data Format at Fall 2024 Allotrope Connect in Boston: Link
- ZONTAL, a SciY company, is making in the Fairdata space: Link
- Enabling Automated End-to-End Chromatographic Data Workflows and Accelerated Data Insights with the Allotrope Simple Model (ASM) Vendor-Neutral Data Format: Link
- Unlock Chromatography Data with FAIR Data Standards: Link1, Link2
- How ASM connectors drive AI/ML to accelerate wet & dry lab workflows by Benchling at Fall 2024 Allotrope Connect in Boston: Link
- Unlocking Scalable Data and AI Power by Scitara at Fall 2024 Allotrope Connect in Boston: Link
- Unlocking the Power of Scientific Data with Scitara DLX Streaming Capabilities: Link1 Link2
- OpenChrom at Fall 2024 Allotrope Connect in Boston: Link
- Lablicate Supports Open Data Standards with Allotrope and OpenChrom: Link
- 2024 Fall Allotrope Connect event is hosted by Thermo Fisher Scientific: Link

AF Community and Events

2024 Fall Allotrope Connect Workshop

Exciting 2024 Fall Allotrope Connect workshop, held on November 19-20, 2024, in Boston, MA, and hosted by Thermo Fisher Scientific, featured a diverse lineup of presentations aimed at advancing data standardization, laboratory innovation, and the integration of digital ecosystems in the chemical, pharmaceutical, and life sciences sectors.

The presentations underscored Allotrope's leadership in driving data standardization, interoperability, and innovation in laboratory environments. They emphasized collaborative efforts to harness data for AI/ML applications and the broader goal of building an open, interconnected scientific ecosystem.

- Slides and video recordings are available at the event page: https://www.allotrope.org/2024-fall-allotrope-connect
- Detailed program, presentations abstracts and speaker's bio can be downloaded here
- View the recordings on YouTube

Below is an overview of the key presentations:

Day 1 Presentations Highlights:

- **Welcome and Strategy:** Janet Cheetham (Allotrope Foundation) opened the event and outlined business, product and partnerships strategies.

- **Opening Keynote:** Richard Milne (Thermo Fisher) emphasized the importance of creating a collaborative digital ecosystem, addressing the need for orchestrated workflows in heterogeneous laboratory environments.

- **ELN Data Archiving with ASM:** Sreeni Yetukuri (Merck) and Spencer Gardiner (ZONTAL) shared a solution for archiving over 30 TB of legacy ELN data into a FAIR format using a draft ASM model, ensuring future usability and compliance.

- **Discoverable Data and Ontologies:** Kirsten Gesenberg (Biovia) explored challenges in creating universal ontologies for laboratory procedures, highlighting complexities in achieving seamless data discoverability.

- **Shaping the Future of ELNs:** Moderated by Vinny Antonucci (Allotrope Foundation), this discussion focused on how Allotrope can enhance ELN effectiveness and drive integration within digital labs.

- **Automating Instrument Data Capture:** Nick Floeck and Milton Yu (Benchling) demonstrated ASM connectors that automate instrument data capture and analytics, accelerating both wet and dry lab workflows.

- **Data Validation and Conversion:** Nathan Clark (Ganymede) presented a flexible validation approach to convert any dataset into Allotrope-compliant formats using schema mapping techniques.

- **Improving Laboratory Efficiency:** Tony Kappen (Agilent) showcased the role of open ecosystems and technology-neutral formats in improving laboratory productivity through AI/ML and automation.

- **Chromatographic Data Standardization:** Chris Siegler (Merck) presented how ASM facilitates automated chromatographic workflows, enabling advanced data consumption and integration with AI/ML pipelines.

- **ADF and ASM Demonstration:** Matthias Mailänder (Lablicate) demonstrated how OpenChrom supports ASM and ADF formats, enabling multi-vendor data integration for chromatography and beyond.

- **Advancing Open-Source Converters:** Vinny Antonucci (Merck) discussed the usage of open-source software, and Nathan Stender (Benchling) showcased the Allotropy open-source library for converting instrument data into ASM-compliant formats.

- **Data Standards for AI/ML:** Kashef Qaadri (Bio-Rad) led a discussion on making data machine-ready for AI/ML by focusing on quality, consistency, and accessibility.

- **Working Groups, Release and Product Update:** presented by Ben Woolford-Lim, the Allotrope Product Team

Day 2 Presentations Highlights:

- **Opening Keynote:** Ryan Snyder (Thermo Fisher) addressed the challenges of rapid data evolution and highlighted Thermo Fisher's efforts in building connected platforms for future-ready labs.

- **Scaling Digital Science:** Bill Goodman (Thermo Fisher) discussed strategies for transitioning from pilot projects to production-scale digital transformations through orchestrated collaboration.

- **Instrument Integration and Knowledge Graphs:** Sai Guttikonda (Merck) shared insights on using Allotrope frameworks to integrate instruments and create knowledge graphs, enhancing lab data flow.

- **Transforming Lab Connectivity:** David Levy (Scitara) presented scalable strategies for unlocking the value of lab data, emphasizing ASM's role and AI's disruptive potential.

- **Streamlining Method Validation:** Dr. Susanne Bauerschmidt (PharmaLex) demonstrated a GAMP5-compliant platform that leverages ASM to automate analytical method validation for regulatory compliance.

- **From Document to Executable Methods:** Dr. Christof Gaenzler (ZONTAL) and Kyle Larsen (USP) showcased USP's journey to digitize monographs into ASM-formatted executable methods, enabling seamless data exchange across systems.

- **ASM Automatic Creation:** Taylor Janoe, Jon Adams, and Mark Spears (Virscidian) presented an automated process for converting raw instrument data into ASM while maintaining workflow context.

- **Ardia Platform:** Stephane Houel (Thermo Fisher) introduced the Ardia Platform, highlighting its role as an open ecosystem for chromatography and mass spectrometry with robust metadata and FAIR compliance.

- **Roadmap and Implementation Strategies:** Janet Cheetham (Allotrope Foundation) led a discussion on the Allotrope product backlog and implementation including federation, partnership, prototyping solutions and market scalability.

Looking Forward

The Allotrope Product Team is looking forward to another productive 1st quarter of 2025. 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