Constructing Allotrope Simple Models (ASM) and Validating with JSON Schema to Achieve Data Interoperability

Allotrope Virtual Connect, Fall 2021
September 2021
The Allotrope Simple Model (ASM)

- ASM is a simple text representation of the Allotrope tabular models using JSON.
- It uses terms from the Allotrope Foundation Ontology (AFO), and it leverages the Allotrope Data Model (ADM) already defined and governed by Allotrope and SMEs.
- **Allotrope tabular models apply to domains where there is a single business object being measured and all measurements directly relate to this object.**
- ASM utilizes *JSON Schema* standard for validation
ASM Lightweight JSON format

- Minimizing verbosity
- Fast parsing
- Intuitive structure
- Cross platform way for representing data
- Term representation and coding is similar.
- Extensive off-the-shelf tooling & support in almost every language (Python, Java, C# etc.)
- Widely used to deliver and exchange data using Representational State Transfer (REST) APIs. (A client/server architecture for a scalable and distributed Web)
- Noted limitation in respect to what objects can be modeled. This limitation prohibits developers in a positive way, making the code simpler, more predictable, and easy to read.
The Allotrope Simple Model (ASM)

- The JSON object format (collection of name/value pairs) aligns very well with the structure of Tabular Model
- The Aggregation Model is an extension of the Tabular Model
- It allows aggregation of data attributes under a unified context and/or an indexed collection of similar data pattern.
- The main characteristics of an Aggregation Model:
  - A set of key/value pairs (Table)
  - Keys are terms in the AFO providing a bridge to semantic usage.
  - The context of the key/value pairs is represented only by the container they are defined in
  - Concepts can be aggregated as:
    - a collection (e.g. a peak list contains peaks).
    - facets (attributes) of another concept (e.g. retention time and description).
- The Modeling WG is using Excel as a tool to create the Aggregation models
AFO, ADM, ASM and ADF Dataflows
AFO, ADM, ASM and ADF Dataflows
Vendor proprietary format

Data Acquisition

Push

Pull

Data

Request

Data

- Pub/Sub
- IoT
- ESB
- etc.

Storage

- Database/Data lake
- Knowledge Graph
- Triple Store
AFO, ADM, ASM and ADF
Dataflows

Vendor proprietary format → AFO terms

Data Acquisition → Mapping proprietary terms to AFT/AFO

- Database/Data lake
- Knowledge Graph
- Triple Store

Vendor proprietary format

Bioportal

<table>
<thead>
<tr>
<th>B12</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
</tbody>
</table>
AFO, ADM, ASM and ADF Dataflows

Vendor proprietary format \rightarrow AFO terms

Data Acquisition \rightarrow Mapping proprietary terms to AFT/AFO

Vendor A
- Measurement ID
- TimeOfMeasure

Vendor B
- Measurement ID
- Time

Vendor C
- Measurement ID

Measurement time

ATM

AFT Preferred Label
- measurement identifier
- http://purl.allotrope.org/ontologies/result#AFR_000121
- http://purl.allotrope.org/ontologies/result#AFR_0000952
- http://purl.allotrope.org/ontologies/result#AFR_0001116

AFO Parameter Allotrope URI
- http://purl.allotrope.org/ontologies/result#AFR_0001118
- http://purl.allotrope.org/ontologies/result#AFR_0001119
- http://purl.allotrope.org/ontologies/result#AFR_0001120
- http://purl.allotrope.org/ontologies/result#AFR_0001587
- http://purl.allotrope.org/ontologies/result#AFR_0001584

- Database/Data lake
- Knowledge Graph
- Triple Store
AFO, ADM, ASM and ADF Dataflows

Vendor proprietary format → Data Acquisition

Dataflows:
- Mapping proprietary terms to AFT/AFO
- Structuring according to ADM

Any AF model, semantically rich (RDF)

Storage:
- Database/Data lake
- Knowledge Graph
- Triple Store

AFO, ADM, ASM and ADF

Dataflows

Vendor proprietary format → Data Acquisition → Mapping proprietary terms to AFT/AFO → Structuring according to ADM → Any AF model, semantically rich (RDF)
AFO, ADM, ASM and ADF Dataflows

Vendor proprietary format → AFO terms → Mapping proprietary terms to AFT/AFO → Structuring according to ADM → Containerizing within ADF

Any AF model, semantically rich (RDF)

- Database/Data lake
- Knowledge Graph
- Triple Store

ADF file
- Data Description
  - Graph Store
- Data Cubes
- Data Package
- Audit Trail
- Check Sums
AFO, ADM, ASM and ADF Dataflows

Vendor proprietary format → Data Acquisition → Mapping proprietary terms to AFT/AFO → Structuring according to ADM → Containerizing within ADF → Storage
- Database/Data lake
- Knowledge Graph
- Triple Store

Any AF model, semantically rich (RDF) → Unpacking the ADF

AFT/AFO

Data Acquisition

Mapping proprietary terms to AFT/AFO

Structuring according to ADM

Containerizing within ADF

Unpacking the ADF

Any AF model, semantically rich (RDF)
AFO, ADM, ASM and ADF Dataflows

Vendor proprietary format → AFO terms

1. Data Acquisition
2. Mapping proprietary terms to AFT/AFO
3. Structuring according to ASM
4. Tabular/Aggregation model (JSON)

Example JSON:

```json
{
  "measurement time": "2015-09-24T03:47:13.001Z",
  "measurement identifier": "413befffd",
  "analyst": "Agent-1",
  "sample identifier": "unknown-10",
  "equipment serial number": "123-AB",
  "batch identifier": "Batch-333",
  "conductivity": {
    "value": "273000",
    "unit": "S/m"
  },
  "temperature": {
    "value": "28.6",
    "unit": "degC"
  }
}
```
AFO, ADM, ASM and ADF Dataflows

Vendor proprietary format → Data Acquisition → Mapping proprietary terms to AFT/AFO → Structuring according to ASM → Container of choice → Storage

- Database/Data lake
- Knowledge Graph
- Triple Store

Tabular/Aggregation model (JSON)
AFO, ADM, ASM and ADF
Dataflows

Cloud

On-premise

- Database/Data lake
- Knowledge Graph
- Triple Store

Vendor outputting ADF natively, any model (RDF)

Unpacking the ADF

ADF file

- Data Description
- Graph Store
- Data Cubes
- Data Package

Data Acquisition

- Storage
  - Database/Data lake
  - Knowledge Graph
  - Triple Store

Vendor outputting ADF natively, any model (RDF)
Vendor outputting ASM natively (Rest interface), Tabular/Aggregation model (JSON)
ASM JSON Model Example (Demo)
REST API and ASM (Demo)
RESTful API and ASM Potential Play in Lab Automation

• REpresentational State Transfer (REST) is an architectural style for providing standards between sw systems on the web.
• It is a mechanism that allows different subsystems used in a laboratory and the organization (e.g. LIMS, ELN, ERP, management systems...) to exchange data in a way which is:
  – standardized
  – structured
  – documented
  – secured
• RESTful APIs are one of the key concepts behind software:
  – integrations
  – interfacing
  – interoperability
• Using RESTful APIs with the ASM data format gives an organization a powerful tool to:
  – automate various functionalities between subsystems
  – push or modify data into subsystems
  – export data from instruments/subsystems
  – integrate with 3rd party application/program/system/analytics used in the lab
Real-time reports with native ASM REST message

Conductivity simulator with ASM REST API
posting in real time ASM Conductivity results into Google sheets

REST API (standard HTTP methods e.g., POST, PUT)

Conductivity Meter

ASM Conductivity result

{ASM}

Payload

Google Sheets

©2021 Allotrope Foundation
Real-time reports with native ASM REST message

Conductivity simulator with ASM REST API
posting ASM Conductivity results into Google sheets
### Conductivity Results Table

<table>
<thead>
<tr>
<th>Comment</th>
<th>$asm.manifest</th>
<th>measurement Identifier</th>
<th>measurement time</th>
<th>analyst</th>
<th>sample Identifier</th>
<th>equipment serial</th>
<th>batch Identifier</th>
<th>conductivity</th>
<th>conductivity.value</th>
<th>conduct temperature</th>
<th>temperature.value</th>
<th>temperature.unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conductivity</td>
<td><a href="http://port.alotropo.org/admin/manifest/con">http://port.alotropo.org/admin/manifest/con</a></td>
<td>288382456 XYZ</td>
<td>8/28/2021 08:13</td>
<td>Angenotts1</td>
<td>unknown-10</td>
<td>278382456 XYZ</td>
<td>22.42536</td>
<td>22.425360785</td>
<td>50m</td>
<td>(&quot;value&quot;: 17)</td>
<td>17</td>
<td>degC</td>
</tr>
<tr>
<td>Conductivity</td>
<td><a href="http://port.alotropo.org/admin/manifest/con">http://port.alotropo.org/admin/manifest/con</a></td>
<td>288382456 XYZ</td>
<td>8/28/2021 08:13</td>
<td>Angenotts1</td>
<td>unknown-10</td>
<td>278382456 XYZ</td>
<td>37.71666</td>
<td>37.716660651</td>
<td>5m</td>
<td>(&quot;value&quot;: 18)</td>
<td>18</td>
<td>degC</td>
</tr>
<tr>
<td>Conductivity</td>
<td><a href="http://port.alotropo.org/admin/manifest/con">http://port.alotropo.org/admin/manifest/con</a></td>
<td>288382456 XYZ</td>
<td>8/28/2021 08:14</td>
<td>Angenotts1</td>
<td>unknown-10</td>
<td>278382456 XYZ</td>
<td>30.460000</td>
<td>30.46000583</td>
<td>5m</td>
<td>(&quot;value&quot;: 19)</td>
<td>19</td>
<td>degC</td>
</tr>
<tr>
<td>Conductivity</td>
<td><a href="http://port.alotropo.org/admin/manifest/con">http://port.alotropo.org/admin/manifest/con</a></td>
<td>288382456 XYZ</td>
<td>8/28/2021 08:15</td>
<td>Angenotts1</td>
<td>unknown-10</td>
<td>278382456 XYZ</td>
<td>17.88419</td>
<td>17.88419771</td>
<td>5m</td>
<td>(&quot;value&quot;: 20)</td>
<td>20</td>
<td>degC</td>
</tr>
<tr>
<td>Conductivity</td>
<td><a href="http://port.alotropo.org/admin/manifest/con">http://port.alotropo.org/admin/manifest/con</a></td>
<td>288382456 XYZ</td>
<td>8/28/2021 08:15</td>
<td>Angenotts1</td>
<td>unknown-10</td>
<td>278382456 XYZ</td>
<td>7.568958</td>
<td>7.5689580056</td>
<td>5m</td>
<td>(&quot;value&quot;: 21)</td>
<td>21</td>
<td>degC</td>
</tr>
<tr>
<td>Conductivity</td>
<td><a href="http://port.alotropo.org/admin/manifest/con">http://port.alotropo.org/admin/manifest/con</a></td>
<td>288382456 XYZ</td>
<td>8/28/2021 08:16</td>
<td>Angenotts1</td>
<td>unknown-10</td>
<td>278382456 XYZ</td>
<td>10.42854</td>
<td>10.428540646</td>
<td>5m</td>
<td>(&quot;value&quot;: 22)</td>
<td>22</td>
<td>degC</td>
</tr>
</tbody>
</table>

---

**Graph:**
- **Conductivity.value and temperature.value**

The shared Google spreadsheet’s table and chart are updated in **real-time** as new ASM conductivity results arrive.
ASM Generation:
ADM to ASM Model Transformation Tool
ADM ↔ ASM Semantic Compatibility

Semantically rich modeling (RDF)

AFT/AFO Terminology

Modeling (ADM)

ADM ↔ ASM Compatibility

ADM to ASM transformation

Tabular ("Simple") modeling (JSON)

Modeling (ASM)

©2021 Allo trope Foundation

ADM and ASM are produced during the governance process
ADM to ASM Transformation

Semantically Rich Model (ADM)

Using SHACL to describe and validate ADM/RDF instance data

Model Consistency: Data and semantic context

ADM to ASM Transformation

Bidirectional model compatibility

Simplified Model (ASM)

Using JSON Schema to describe and validate ASM/JSON instance data

Model Validation

JSON Schema

{JSON}
ASM Validation
In Action
(Demo)
ASM JSON Schema Validation

- **JSON schema** validator evaluate conformance of the **ASM Data** instance with the set of conditions and rules in the associated **JSON schema**.
- A list of many off-the-shelf validators, written in different languages, is available at [https://json-schema.org/implementations.html](https://json-schema.org/implementations.html)
A set of pairs: **Self described “Unique Information Parameter”** and the **“Value/Unit”**

“Unique Information Parameter” : “Value/Unit”
Thanks for your attention!

Allotrope Foundation Product Team

- Ben Woolford benjamin.woolford-lim@allotrope.org
- Amnon Ptashek amnon.ptashek@allotrope.org