Status of LC-UV Data Model

November 2018
Agenda

- Scope
- Workflow
- Submission
- Patterns
- What can you do with it?
- Future work
Initial Release Use-Cases

• As an analytical chemist I need to be able to retrieve my HPLC-UV analytical method parameters from a standardized file format.
  – Confirm the correct method parameters were used in obtaining the data
  – Data Mining to be able to predict/prioritize effective conditions for method development

• As an analytical chemist I need to retrieve standard HPLC-UV results and instrument descriptions.
  – Automated Result Reporting
  – Data Mining for Lot Trend analysis, control chart analysis of instrument / method performance.
Full Graph Development Workflow

Initiation | Development Cycle | Finalize | Review | Release

Promoters
- Assemble Core Team
- Determine Scope

Scientific SMEs
- Scientific Review
- Model and Term Development

Core Team
- Semantic Review

Semantic SMEs

Public Review
- SPARQL queries
- SHACL rules

Instance data conversion

ADM*
- AFO update incl. testing
- ADM* package incl new patterns
- Documentation
What’s in a submission?

Required:
- Class model diagrams (CMAP)
  - Visual documentation of the proposed models, including all entities, relationships, and cardinalities
- New term definitions (Excel)
  - Listing of the new semantic terms required to be added to AFO
- Example test data (native and TTL)
  - Example dataset in a native format and a TTL version of the same dataset conforming to the proposed model
- SPARQL queries (RQ)
  - Example SPARQL queries i.e. how to ask questions of the proposed model.

Optional:
- Automated SHACL rules (SHACL/TTL)
  - SHACL implementation of the proposed model to automatically verify that data matches the model
- Pattern model diagrams (CMAP)
  - Visual documentation of any new patterns defined by the model that should be reused in development of other models.
GitLab

• Everything hosted on GitLab repository
• Link is https://gitlab.com/allotrope-review/lc-uv
• Please raise any issues or comments on the issue tracker there
## Allotrope LC-UV Model Review

**Project ID:** 8133889

### Repository Details
- Star: 0
- Fork: 0
- SSH: `git@gitlab.com:alltrope-lc-uv`

### Branches and Tags
- Master
- `lc-uv`

**Merge branch 'adf-sample' into 'master'**
Helge Krieg authored 2 weeks ago

### Files
<table>
<thead>
<tr>
<th>Name</th>
<th>Last commit</th>
<th>Last update</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>diagrams</code></td>
<td>Upload New File</td>
<td>1 month ago</td>
</tr>
<tr>
<td><code>miscellaneous</code></td>
<td>renamed folder</td>
<td>1 month ago</td>
</tr>
<tr>
<td><code>pattern</code></td>
<td>updated release notes, renamed file to be consistent, added pa...</td>
<td>1 month ago</td>
</tr>
<tr>
<td><code>query</code></td>
<td>Upload New File</td>
<td>1 month ago</td>
</tr>
<tr>
<td><code>test data</code></td>
<td>Added example ADF based on 002-1401.D dataset with data c...</td>
<td>3 weeks ago</td>
</tr>
<tr>
<td><code>validation</code></td>
<td>Upload New File</td>
<td>1 month ago</td>
</tr>
<tr>
<td><code>vocabulary</code></td>
<td>Upload New File</td>
<td>1 month ago</td>
</tr>
<tr>
<td><code>.gitignore</code></td>
<td>initial commit</td>
<td>2 months ago</td>
</tr>
<tr>
<td>README.md</td>
<td>Update README.md</td>
<td>1 month ago</td>
</tr>
</tbody>
</table>
**Upload New File**
Wes Schafer authored 1 month ago

<table>
<thead>
<tr>
<th>Name</th>
<th>Last commit</th>
<th>Last update</th>
</tr>
</thead>
<tbody>
<tr>
<td>..</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instance_Graph__002-1401.D.cmap</td>
<td>Upload New File</td>
<td>1 month ago</td>
</tr>
<tr>
<td>LCUV_Model__Class_Model.cmap</td>
<td>Upload New File</td>
<td>1 month ago</td>
</tr>
<tr>
<td>README.md</td>
<td>added folder structure</td>
<td>2 months ago</td>
</tr>
</tbody>
</table>

**Diagrams**

This folder contains diagrams about model of LC-UV.

**How to open .cmap files?**

CMAP files were created with Conceptual Mapping Tool (CMAP Tools). You can [download CMAP Tools here](#). Installation in Windows runs with standard user privileges.
<table>
<thead>
<tr>
<th>Name</th>
<th>Last commit</th>
<th>Last update</th>
</tr>
</thead>
<tbody>
<tr>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADFRuns.rq</td>
<td>Update ADFRuns.rq</td>
<td>1 month ago</td>
</tr>
<tr>
<td>ExpectedResults_002-1401</td>
<td>Upload New File</td>
<td>1 month ago</td>
</tr>
<tr>
<td>README.md</td>
<td>Update README.md</td>
<td>1 month ago</td>
</tr>
<tr>
<td>acquisitionStart.rq</td>
<td>Update acquisitionStart.rq</td>
<td>1 month ago</td>
</tr>
<tr>
<td>materials.rq</td>
<td>Update materials.rq</td>
<td>1 month ago</td>
</tr>
<tr>
<td>operatingSettings.rq</td>
<td>Update operatingSettings.rq</td>
<td>1 month ago</td>
</tr>
<tr>
<td>peakFacets.rq</td>
<td>Update peakFacets.rq</td>
<td>1 month ago</td>
</tr>
<tr>
<td>processParticipants.rq</td>
<td>Update processParticipants.rq</td>
<td>1 month ago</td>
</tr>
</tbody>
</table>

**Query**

This folder contains SPARQL queries based on competency questions in order to query data from LC-UV model.

Software such as Apache Jena and Protege can store graph instance data and perform SPARQL queries against it.

They are currently configured to work from an Apache Jena Fuseki dataset named "LC-UV" with two rdf graphs, one containing the instance data from this submission "002-1401" and a graph "Merged-ontologies" that contains the proposed AFO owl ontology in this submission.
Table:

<table>
<thead>
<tr>
<th>Name</th>
<th>Last commit</th>
<th>Last update</th>
</tr>
</thead>
<tbody>
<tr>
<td>source data</td>
<td>Upload New File</td>
<td>1 month ago</td>
</tr>
<tr>
<td>002-1401.D.adf</td>
<td>Added example ADF based on 002-1401.D dataset...</td>
<td>3 weeks ago</td>
</tr>
<tr>
<td>ADM_Instance_Data_002-1401.D.ttl</td>
<td>Update ADM_Instance_Data_002-1401.D.ttl.txt</td>
<td>1 month ago</td>
</tr>
<tr>
<td>README.md</td>
<td>added folder structure</td>
<td>2 months ago</td>
</tr>
</tbody>
</table>

**Test data**

This folder contains test data based on LC-UV model. These are mainly examples of instance data serialized as Turtle files.
### Vocabulary (Allotrope Review)

This folder contains a merged artifact of candidate recommendation of the Allotrope Foundation Ontologies (AFO) for the domain of LC-UV, version CR/2018/09/17. This file was copied from merged AFO version CR/2018/09/17.

You can find all individual artifacts and all versions of AFO at [http://purl.allotrope.org](http://purl.allotrope.org)

### How to open ontologies for review?

We recommend using Protégé for viewing ontologies.

There are different options for getting access to AFO:

- [OntologyLib](http://ontologlib/)
- [AFO Reviewer](http://afo-reviewer/)
Patterns

- Part of the process of modelling involves identifying reusable patterns.
- These patterns should be small subgraphs that have a consistent shape throughout.
- Patterns may be specific to a domain, e.g. the pattern for an LC-UV system, or apply across domains, e.g. the pattern for a piece of equipment.
- Development of patterns and their use ensures consistency across different datasets and domains.
What can we do with it?

Answer key questions about your experiments through SPARQL queries

- Who?
  - Who submitted the sample?
- What?
  - What sample was submitted?
- Where?
  - Where was the experiment performed?
- When?
  - When was the sample processed?
- How?
  - How was the experiment performed?
- Why?
  - What was the purpose of the experiment?
## On-Going Efforts

<table>
<thead>
<tr>
<th></th>
<th>Initial SME data collection</th>
<th>Initial ADM / CMAP</th>
<th>Technical Review</th>
<th>Completion of submission</th>
<th>Public Review</th>
<th>Release</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial LC -UV</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Release funded / Expected EOY/1Q</td>
</tr>
<tr>
<td><strong>GC</strong></td>
<td>Funded</td>
<td>Expected EOY</td>
<td>Expected 1Q2019</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Columns</strong></td>
<td>Funded</td>
<td>Expected EOY</td>
<td>Expected 1Q2019</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Calibrated results</strong></td>
<td>v 1.1.5 CMAP avail</td>
<td>Funded</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LC-MS (SQD)</strong></td>
<td>v 1.1.5 CMAP avail</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Non-chromatography full graph models: DSC (in-progress), XRPD, TGA (planned)
Next steps

• Try it!
  – APN members can integrate this model into their products and give feedback for improvement
  – Allotrope can provide targeted support and maintenance for APN members building LC-UV converters using ADM

• Test consistency of model across implementations
  – Once multiple converters are available to generate LC-UV ADFs, the consistency in ADM application will be assessed and adjustments made if warranted

• Extend the data model suite
  – Include additional domains needed to support workflows (materials, specifications, analytical procedures, etc)