Methods Database PoC at Merck and GSK

Sept 15, 2021
Agenda

- Project Summary
- PoC Scenarios
- Current State at Agilent and ZONTAL
- Next Steps at Merck and GSK
- Roadmap
Data reproducibility continues to be a costly problem in life sciences. One study estimates that of the $56B spent on pre-clinical research each year, half (approx. $28B) is not reproducible\(^1\).

Current analytical methods used in the Pharmaceutical Industry face challenges of interoperability and reproducibility due to:
- Manually kept and managed (often paper based)
- Stored in proprietary software platforms

Improving both automation and interoperability of analytical methods may:
- Improve quality
- Optimize efficiency and
- Reduce cost

By...
- Improving ability to effectively transfer methods
- Enabling Digital b2b Methods transfer (e.g., from CRO to Pharma)
- Reducing or eliminating the need to manually enter data from paper methods into electronic platforms
- Enhancing reproducibility and providing full content of the experimentation
- Improve Out of Spec (OOS) analysis capability and potentially leverage data standards for machine-learning

\(^1\) [https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002165](https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002165)
Ontologies provide an unconstrained vocabulary we can use to describe things (instances) in our open world and give them a meaning.

Data structures (schemas, templates) describe how to use the ontologies for a given purpose (reproducible, predictable, verifiable) way.

Shapes Constraint Language (SHACL expressed as RDF) is a WC3 standard to do this, used for ADM.

Phase 1: An interoperable data standard for Chromatography Data Systems

Commonly used CDS Parameters

- Flow rate
- Cooler Temp
- Injection Volume
- Valve position
- Full scan option
- Range
- Valve position at time
- Full scan (step)
- Pressure limits
- Solvent lines
- Solvent % at time

Ontologies provide an unconstrained vocabulary we can use to describe things (instances) in our open world and give them a meaning.

Data structures (schemas, templates) describe how to use the ontologies for a given purpose (reproducible, predictable, verifiable) way.

Shapes Constraint Language (SHACL expressed as RDF) is a WC3 standard to do this, used for ADM.
Phase 2: Creation of CDS Adapters and database to host methods

1) Method Metadata to link to your ELN or LIMS system

2) Column Methods including consistent column descriptions

3) Sequence Parameter to deal with a whole run rather than single injections

Waters Empower CDS Adapter

Agilent Openlab CDS

Methods DB – powered by
Methods Hub Roadmap for closing the Data Loop

LIMS / ELN / Workflow

Samples → Method → Results → Context → Insights

Instrument Control Software

Methods Database

Advanced Analytics

Monographs, Publications

Result Data Adapter for CDS (end ‘21)

Use of AI to monitor Analytical performance

NLP supported import of text-based method information

Closed Loop for sampling and OOS analysis
The Pistoia Alliance’s Methods Hub is envisioned to be a **platform** where semantically **interoperable analytical methods** and supportive tools are available to the Pharmaceutical Industry.

**Methods Database**
- Pistoia alliance and Allotrope foundation
- Standardized HPLC parameters in Allotrope Data Format (ADF). Electronic transfer of analytical methods and results between CDS instruments

**Chromatography Column Data Model**
- Alignment with US Pharmacopeia column metadata

**Methods on Demand**
- Transcription of text-based methods to parameters via NLP into a cloud-based platform with machine readable metadata
OpenLab CDS Integration

PoC Setup 1

Notes:
- OpenLAB CDS Workstation uses local storage
- Use Post-Run Macro or ZONTAL File Watcher to automate Method Upload
OpenLab CDS Integration
PoC Setup 2

Notes:
- Sample Scheduler requires OpenLAB CDS Server
- Use Post-Run Macro or ZONTAL File Watcher to automate Method and Results Upload
Empower Integration

PoC Setup

Support Agreement

Beta License Agreement

ORBIS Public Cloud

ORBIS Portal

ORBIS Gateway

Empower Server

HPLC Agilent 12xx

Run Sequence

Method

Results

Sample Sequence

Results

ZONTAL Public Cloud

OR

AWS Private Cloud

LIMS/ELN

Beta License Agreement

Support Agreement
Live Demo

Agilent already presented ADF method export / import in Fall 2020

What happened since then:

- ADF method graph according to lc-uv-method\REC\2020\09
- Rewrote Agilent LC ADF method export / import code from scratch
- Added pump gradient table
- Support now > 90 % of the LC parameters for a standard Agilent LC
  (Agilent specific taxonomy for method parameters)
Live Demo

**Method**

- **Autosampler**
  - Analog Output: 2.0 V
  - Margin for Negative Absorbance: -1000 mAU
  - Real Detection Aabsorb.: 0.0005 mAU
- **Diode Array Detector**
  - Detector Sampling Rate: 1.0 Hz
  - Ultraviolet Absorbance Measurement: 1
    - Electronic Absorbance Bandwidth: 1.0 to 4.0 nm
    - Electronic Absorbance Wavelength: 250 nm
  - Referring
    - Electronic Absorbance Reference Bandwidth: 1.00 nm
    - Electronic Absorbance Reference Wavelength: 250 nm
- **Analog Output**
  - Analog Output: 1
  - Margin for Negative Absorbance: -1000 mAU
  - Real Detection Aabsorb.: 0.0005 mAU
  - Detector Offset: 5.0%
- **Precondition Evaluation**
  - Ultraviolet Spectrophotometry
    - UV Spectral Absorbance
- **Injection**
  - Ultraviolet Absorbance Measurement
    - Electromagnetic Radiation Chromatogram
      - Description: 1000 A, Wavelength=250 nm
      - Integration
        - Peak List

**Chromatogram**

- Absorbance (mAU)
- Time (s)

---

**Process**

- Ultraviolet Absorbance Measurement
  - Electromagnetic Radiation Chromatogram
    - Description: 1000 A, Wavelength=250 nm
    - Integration
      - Peak List

---

**Legend**

- Green Area: Injection
- Red Area: Electromagnetic Radiation
- Yellow Area: Ultraviolet Absorbance
Next Steps at Merck and GSK

- Integrate the Agilent PoC Configuration into Merck and GSK laboratories
- Automate Integration of OpenLAB CDS to ZONTAL
- Deploy Empower Adapter for Methods and Results
Methods on Demand, Pistoia Alliance

Business Challenge
- Method descriptions are still mainly text-based documents
- Reproducibility of methods limited by interpretation of free text
- Descriptions use different terminology and levels of detail
- Version control often difficult because of number of copies

Objectives
- Parse methods from public and private monographs and academic publications (Natural Language Processing, NLP)
- Extension of method model from instrument method to complete analytical method

Contributors
- Pharma Member Companies, USP (US Pharmacopeia), BP (British Pharmacopeia), CAS, a division of the American Chemical Society (MethodsNow), Elsevier, Pistoia, OSTHUS, ZONTAL
Digitalization of Analytical Methods and SOPs

Document Scanning and OCR

Document Ingestion and NLP
Any Questions? Contact us!