



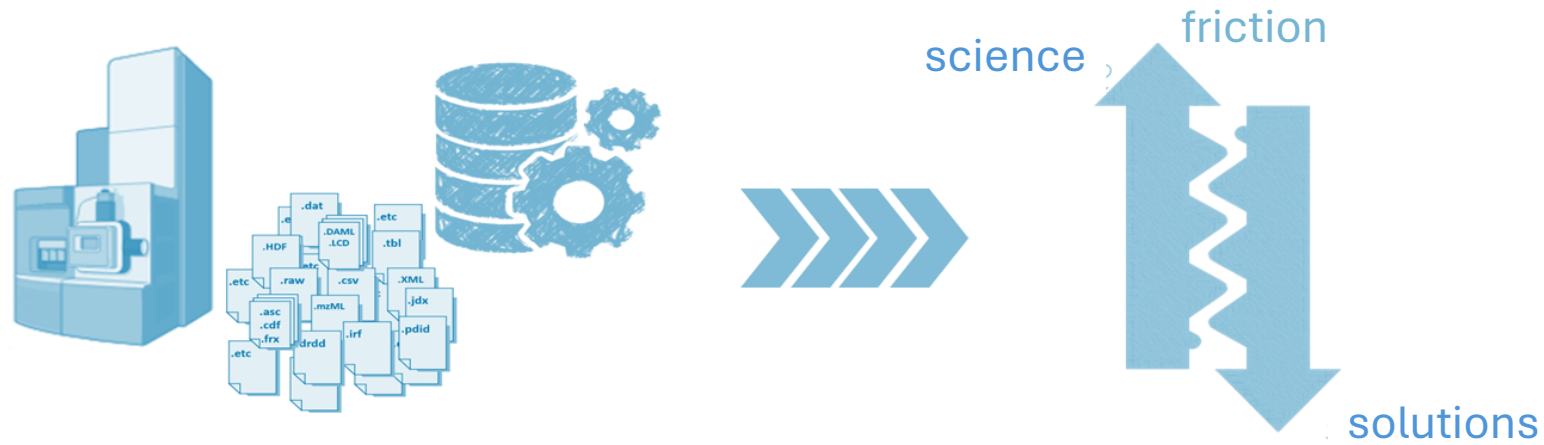
Enabling Automated End-to-End Chromatographic Data Workflows and Accelerated Data Insights with the Allotrope Simple Model (ASM) Vendor-Neutral Data Format

W. Christopher Siegler, Jonathan Fine, Wes Schafer, Steve Borst, Rohit Kulkarni, Pankaj Aggarwal, Vincent Antonucci
(Merck & Co., Inc., Rahway, NJ, USA)

Heiko Fessenmayr, Robert Williams
(Agilent Technologies, Inc.)

Allotrope Connect – November 2024

Problem Statement



- Experiment & measurement data are heterogeneous and not well connected
- Creates tremendous friction executing E2E workflows (capture to consumption)
- Limits (re)use of data, which limits the value created from this key asset
- Heterogeneity leads to friction for users and increased costs, which slows science

2024

Unfortunately, this is still the reality in many R&D laboratories



The Total Cost of Ownership (TCO) of solutions is too high



New solution implementation & integration timelines are too long



Incompatibility with existing technology may limit solution options



Laboratory operations are inefficient, with many routine or manual tasks



Scientists don't have on demand access to institutional knowledge



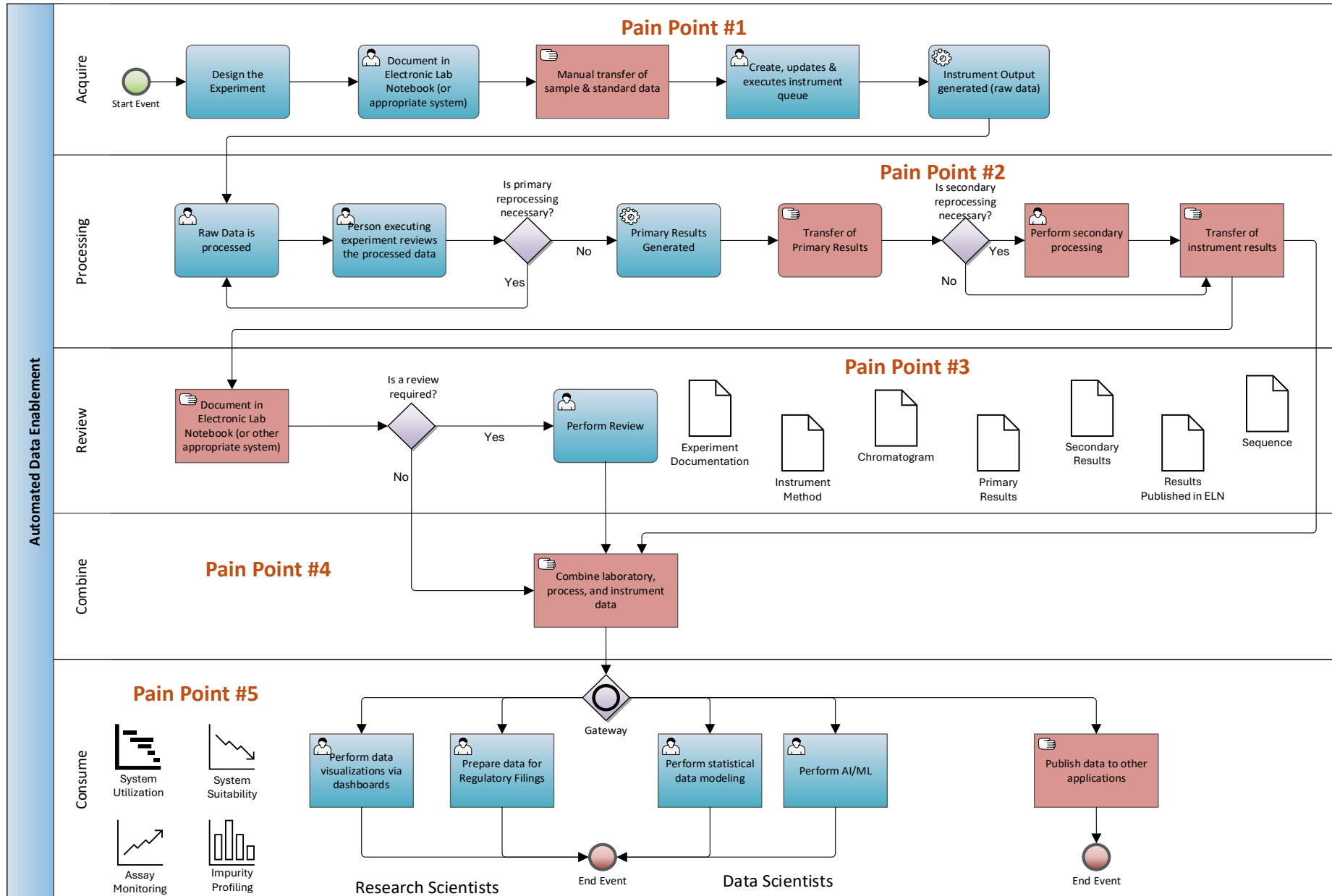
Our greatest asset (creative people) don't have enough time to innovate



Foundations needed for digital transformation are not consistently in place yet to enable more rapid innovation.



Current Business Process



Pain Points:

#1. Manual transfer of experiment data necessary to manually setup and start instrument data acquisition.

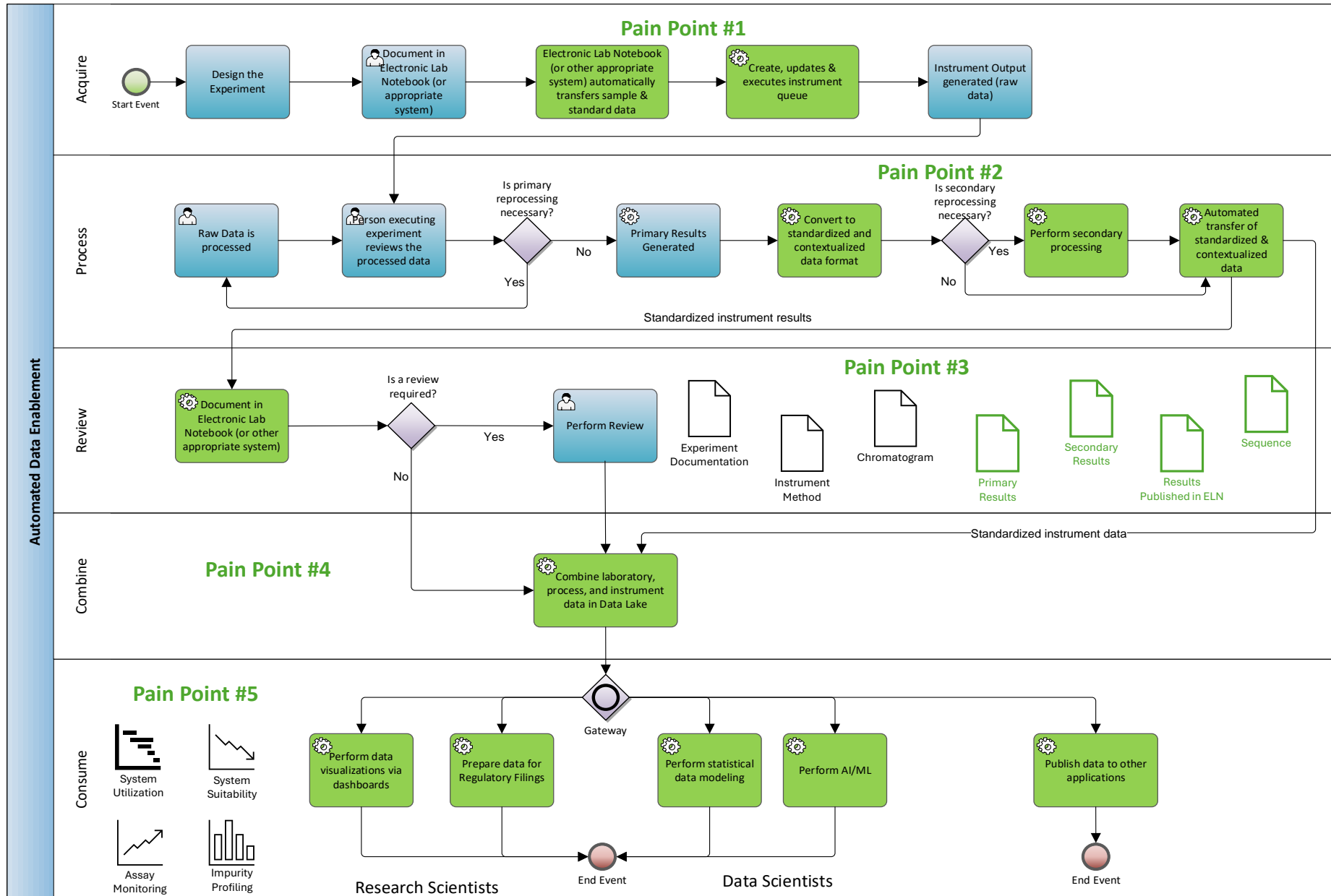
#2. Data in multiple vendor unique formats & secondary data processing occurs in user unique ways.

#3. Manual data transfers and calculations *always* require review.

#4. Manual aggregation of data from disparate sources is time consuming.

#5. Non-structured data adds unnecessary complication for users to consume & publish their data.

Target State Business Process



Pain Points:

#1. Resolved, Sample sequence automated transfer to instrument software.

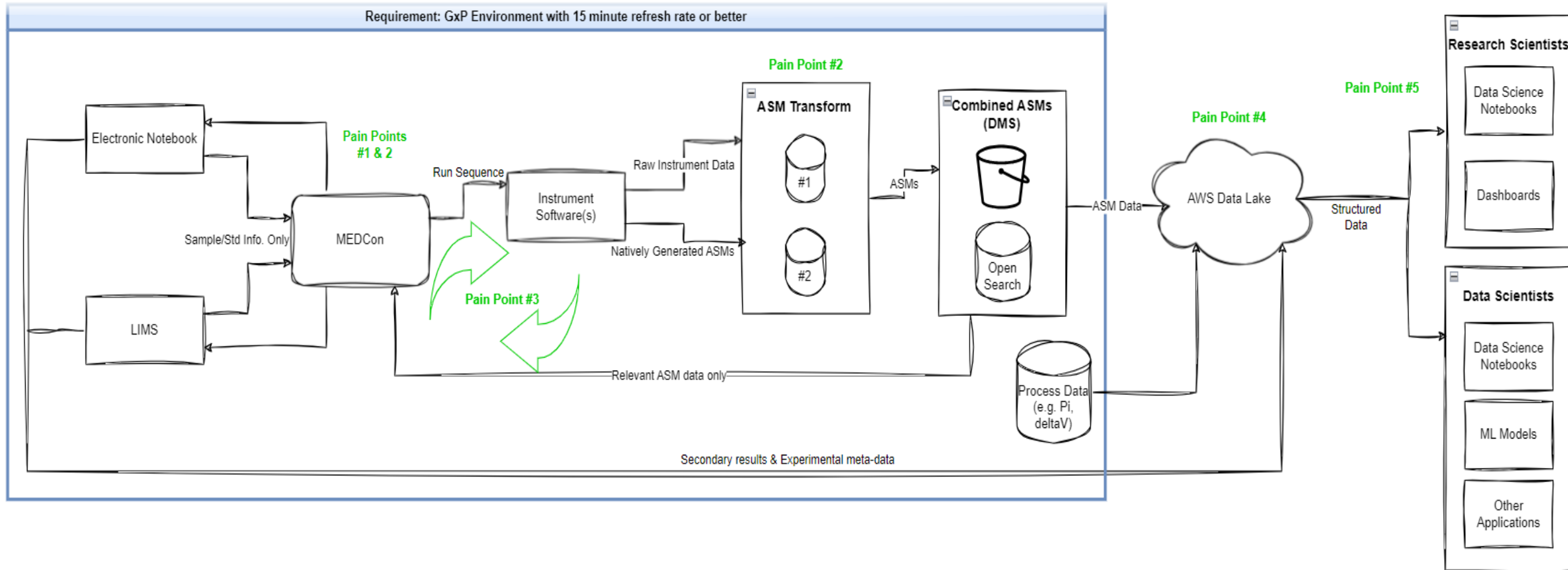
#2. Resolved, Standardized instrument data & secondary calculations.

#3. Reduced, Fewer reviews due to automated transfers.

#4. Resolved, Aggregated data available as data products in data lake.

#5. Structured data in a centralized location simplifies consumption & publication of data by all users.

Target State Conceptual Architecture



Spotlight: Why is ASM Critical to Solution ?

Pain Points

- Numerous manual data transfer pain points by Merck scientists
- Difficulty aggregating data across Empower, Chemstation, Unicorn, etc. systems

ASM for Pain Relief

- Standardized & harmonized format across instrument vendors
 - Simplification of data connectors/pipelines
- Ontologies provide contextualization for data lake aggregation
 - Easier integration with other data products
 - Quicker dataset creation by Data Scientists
 - Necessity for Predictive Modeling, ML, & GenAI applications
- JSON data type is a widely preferred format

Why Use the Allotrope Framework ?



Developed and sustained by an active **community** of experts who co-create data standards and collaborate to apply them to use cases across the life sciences

Robust Operations



Governance



QA & Release Automation



Tooling



Product Team

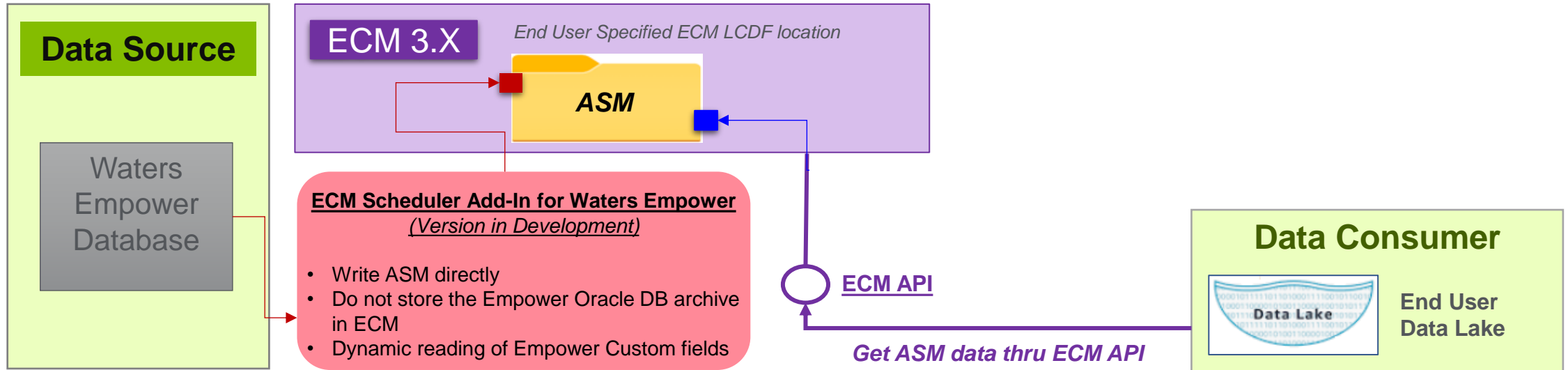


Quarterly Delivery

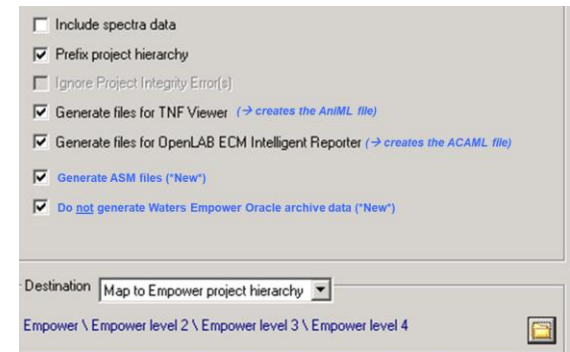


Automated Waters Empower Database Export (ECM 3.X)

Customer Development Project for Exporting Waters Empower Data



New Waters Empower result sets are detected by the ECM Scheduler, automatically converted to ASM objects, and stored in ECM 3.6. The resulting ASM objects can then be extracted from ECM into the End User Data Lake using the ECM API.



“This presentation contains forward-looking statements (including, without limitation, information and future guidance on the company’s goals, priorities, growth opportunities, customer service and innovation plans, new product introductions, financial condition and considerations, and the continued strengths and expected growth of the markets the company sells into, operations) that involve risks and uncertainties that could cause results of Agilent to differ materially from management’s current expectations. The words “anticipate,” “plan,” “estimate,” “expect,” “intend,” “will,” “should” “forecast” “project” and similar expressions, as they relate to the company, are intended to identify forward-looking statements.”

Spotlight: GxP Publication of ASM Results

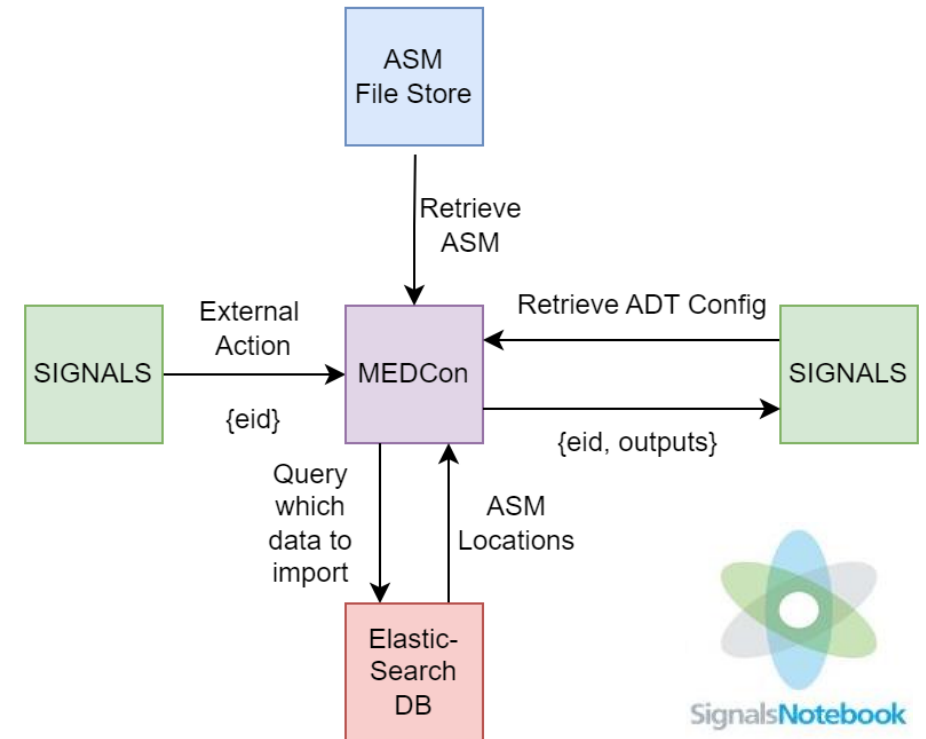
Merck Extensible Data Connector (MEDCon)

What is it?

- In-house built tool
- Enables publication of chromatographic ASM data to Signals ADT
- Allows use of custom Python scripts for complex secondary result calculations
- GxP compliant

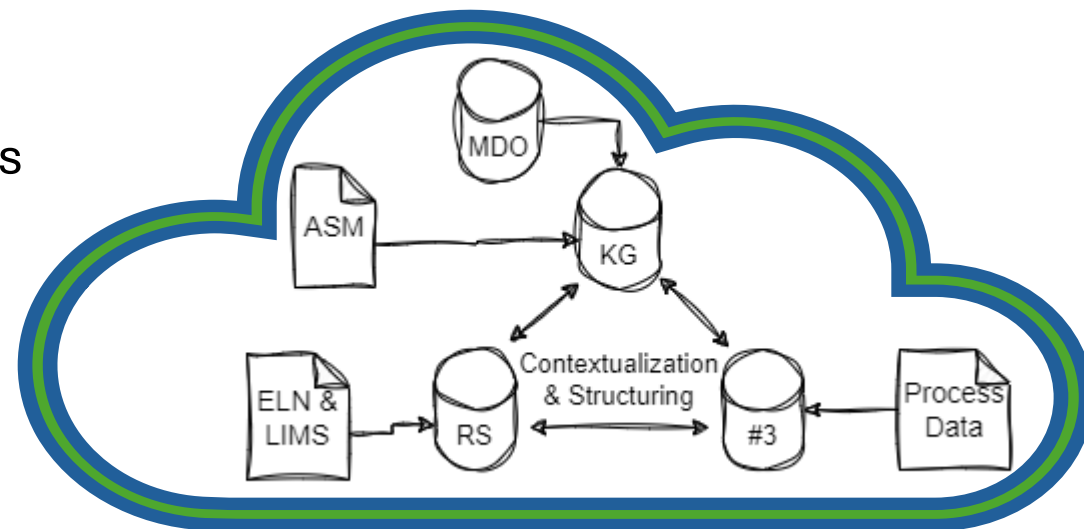
Why is it important?

- Mostly automated transfer of Empower results
 - Minimal user intervention
- Fully validated workflow
 - Reduced SSR/SPR efforts

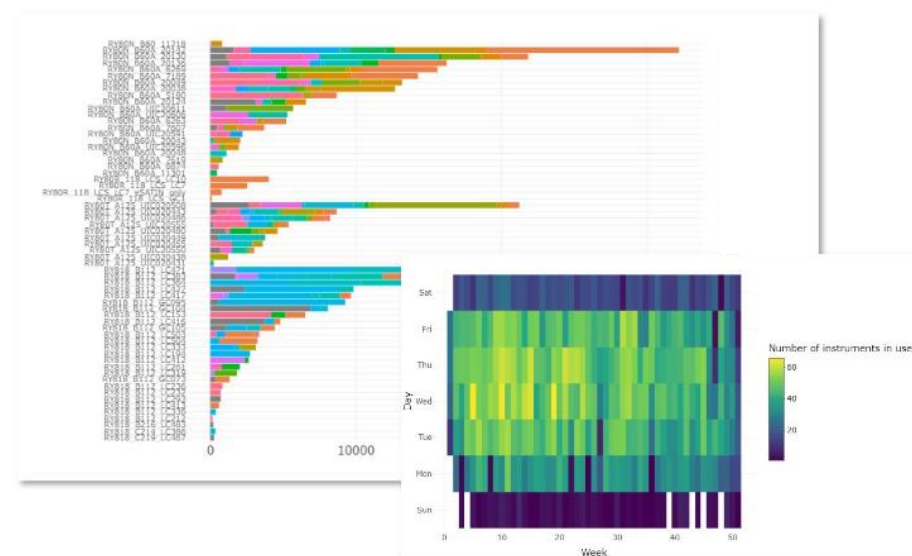
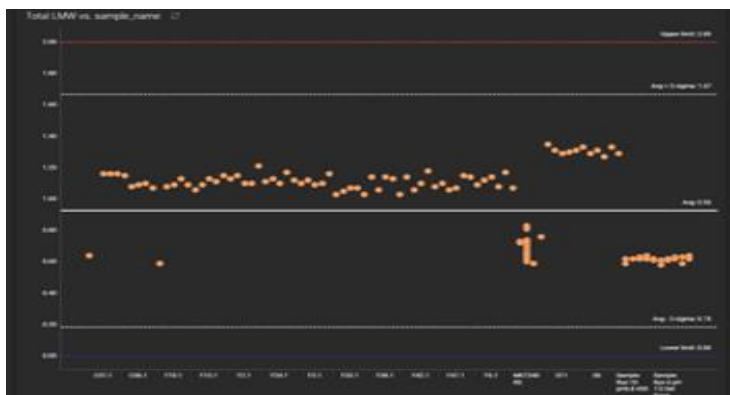


Spotlight: Consumption of ASM Files

- Simplified via standardized & structured data products
 - Centrally located within Merck's AWS Data Lake
 - Knowledge graph (KG) for ingestion of ASM data
 - Redshift (RS) for ingestion of ELN & LIMS data
 - Contextualization via Merck & Allotrope ontologies (MDO)



- Data visualization via Spotfire Dashboards
 - Instrument Utilization
 - Impurity Tracking
 - Assay Monitoring
 - System Suitability
 - CRO Empower Analytics



MEDCon Demo: ASM Result Publication to Signals

Microsoft Teams

Meeting

2024-10-30 20:53 UTC

Recorded by

Fine, Jonathan

Organized by

Fine, Jonathan

System Utilization Dashboard Demo

Microsoft Teams

Meeting

2024-10-30 21:02 UTC

Recorded by

Fine, Jonathan

Organized by

Fine, Jonathan