

Beyond JSON: Centralizing, Modeling, and Aggregating Instrument Data to Enable Deeper Analysis

A case study of Dotmatics Luma automating data acquisition, busting data silos, and enabling complex analytical pipelines at a large biopharma



Background

Across the biopharma industry, R&D teams are struggling to effectively manage their data so that it can be best put to use. The first obstacle is collecting and standardizing huge volumes of data being output in different, often vendor-specific, formats from the variety instruments needed to conduct experiments. This is followed by an even bigger challenge: modifying and aggregating that diverse data output so it can be correlated with other relevant data for deeper analysis. These struggles are amplified in R&D environments where there are dynamic requirements stemming from highly diverse research projects and varied experiment types.

A tempting quick fix is converting output data into a more standardized JSON format; but, this does not appropriately address the true scope of the challenge and fails to facilitate complex downstream analyses.

One major multinational pharmaceutical company saw the shortcomings of this quick-fix approach after a few unsuccessful vendor engagements. The company eventually found a more comprehensive solution with Dotmatics Luma, a groundbreaking R&D data-management platform that enables data acquisition, processing, and deeper contextualization and analysis at scale. The company is not only saving massive amounts of time by connecting thousands of instruments to Luma, but it's also positioning its researchers to more easily perform the complex modeling and analysis needed to better understand their diverse R&D data and drive innovation.

CHALLENGE

For a large biopharma's R&D teams to enable broader data contextualization and advanced analyses, output from diverse instruments used for varied experimentation needed to be standardized, modeled, and integrated to a deeper degree than possible with JSON conversion.

SOLUTION

Dotmatics Luma automates data acquisition from thousands of instruments and uses its advanced data parsing, modeling, integration, and dataflow capabilities to facilitate the complex analysis pipelines the biopharma's teams rely on in their wide-ranging R&D.

RESULTS

Rapid deployment:

- 2000+ instruments connected to Luma within six months (4X faster than expected)

Diverse fleet and experiment support:

- Luma's advanced models accommodate broad instrument diversity and experiment variety

Improved analysis pipelines:

- Luma's flexible data model standardizes diverse data, enabling clean data structures for analysis at any scale
- Instrument output is harmonized with other relevant R&D data for contextualization and analysis in nested pipelines

Challenge

R&D leaders at this major biopharma were well aware of a common shortfall within the industry — most companies' data processing efforts focused on converting output to JSON format, but then stopped there. They wanted to do more. They recognized the true scope of the challenge and sought a more comprehensive solution, one that would be genuinely effective in facilitating complex data processing and analysis, both at scale and across all their diverse programs.

Biopharma's Top R&D Data Processing Requirements

The company's R&D leaders set out to find a vendor whose technology would enable all researchers, working across highly varied projects, to:

1. Acquire outputs from diverse instrument sources.
2. Integrate that data into the appropriate context, at the necessary scale.
3. Facilitate analyses, augmented by sophisticated algorithms.
4. Merge these analyses into subsequent ones, forming nested pipelines.
5. Do all of this without requiring intervention of customer services or project teams.

Despite these clearly defined objectives, team leaders struggled to find a vendor to meet all their needs. They initially engaged with a vendor who helped with the first step in the process, JSON standardization; but this vendor then told the team that addressing the subsequent requirements would necessitate engaging in a costly custom project with a special delivery team.

Underwhelmed, the team continued its search. Just as team leaders were losing hope of finding a vendor who could help them to the degree desired, they engaged with Dotmatics. Team leaders were stunned by how closely the features of Dotmatics' Luma R&D data-management platform matched their needs.

Solution

Within ten minutes of seeing Luma at work, team leaders realized it was the all-encompassing solution they needed to help solve their entire problem. The team felt Luma was categorically different than other solutions on the market because of its:

- seamless linkage of the output processing with the extraordinarily flexible, yet well-governed, data model,
- advanced dataflow and data analysis capabilities, and
- quick deployability.

The team embarked on rigorous proof of concept, which alone accomplished more than multiple years of engagement with its prior vendor. Team leaders quickly decided to forge ahead and roll out Dotmatics Luma in a phased approach.

“We didn't realize this option existed in the wild!”

— R&D team leaders at a larger and growing biopharma

Results

The initial deployment of Dotmatics Luma has significantly exceeded the team's expectations.

Volume: Thousands of Instruments Connected

The team has several thousand instruments to connect. Luma is being deployed much more quickly than anticipated.





- **Goal:** Deploy Luma to 1000 instruments within a year
- **Reality:** Much faster rollout than expected!
 - 1000 instruments in four months
 - 2000 instrument in six months

Currently, about a quarter of the team's instruments have been connected to nested models.

Diversity: Broad Instrument and Experiment Support

With Luma now deployed to more than 2000 instruments, it's supporting the diversity of the team's fleet and the wide variety of experiments conducted. Luma's ability to handle subtle differences in experimental contexts and output types has been essential to this, as has its ability to ingest other data at scale, in exactly the right structure needed.

The following table summarizes key benefits the Luma has delivered:

Previous Challenge	Luma Flow Cytometry Workflow
Inefficient data acquisition and annotation Data collection from numerous systems and instruments meant juggling different upload scripts and often manually annotating and mapping data.	 Connected lab with automatic data upload and intelligent tagging Instrument and system data is automatically uploaded, parsed, tagged, and directed to the appropriate location using Luma's advanced algorithms.
Unmet fleet and experimentation needs Managing the collection and processing of data pouring out of the huge fleet of varied instruments used across a diverse range of research projects was a challenge that demanded a solution bigger than simple file conversion and standardization.	 Broad instrument and experiment support Luma handles subtle differences in experimental contexts and output types with ease; it can ingest data at scale from 100+ types of instruments using out-of-the-box parsers that require no additional configuration, and it can then prepare that data for deeper analysis using flexible data models.
Insufficient JSON standardization Other vendors' solutions stopped with JSON conversion of output data, requiring time-consuming and costly engagements to further ready and progress data through advanced analyses.	 True data transformation and enrichment Output data is processed with an extraordinarily flexible, yet well-governed, data model, fully transforming data into structured representations ready for deeper analysis. Luma can augment and correlate instrument data with other relevant data from external systems, supporting broader contextualization.
Cumbersome analysis process Solutions that stop at JSON conversion force researchers to spend an inordinate amount of time gathering, collating, correlating and otherwise preparing data for the specialized analyses needed within their specific areas of research.	 Advanced analytical pipelines Luma can seamlessly stream ready-to-analyze data into nested analytical pipelines for deeper analysis with sophisticated algorithms, and it can integrate directly with leading scientific analysis tools to more easily facilitate high-value workflows.

