Lab interoperability with DLX using ASM-JSON format

WHITEPAPER:
LAB INTEROPERABILITY WITH DLX

scitara
meet the modern lab™

Scientists across the world find a common language of science. They've created vaccines, medical devices, or other lifesaving products. They work in many different industries, such as pharmaceutical, food and flavor, or environmental.

Yet the technology that has been introduced to scientists to automate and simplify collecting data does not "talk" to one another, and when it does, it doesn't speak a common language. With the increasing number of software, instruments, and devices being utilized in the lab, the amount of data has grown exponentially, meaning more resources are required to collate and interpret data.

In addition, the current silos and fragmented data infrastructure make streamlining the laboratory workflows challenging.

Labs are often supported by data infrastructure which is too application centric.

Each application has its own data interface with a related schema and terminology. Data should be perceived as the most critical and vital asset in support of applications.

> "What science strives for is an utmost acuteness and clarity of concepts as regards their mutual relation and their correspondence to sensory data."

Einstein, "The Common Language of Science."

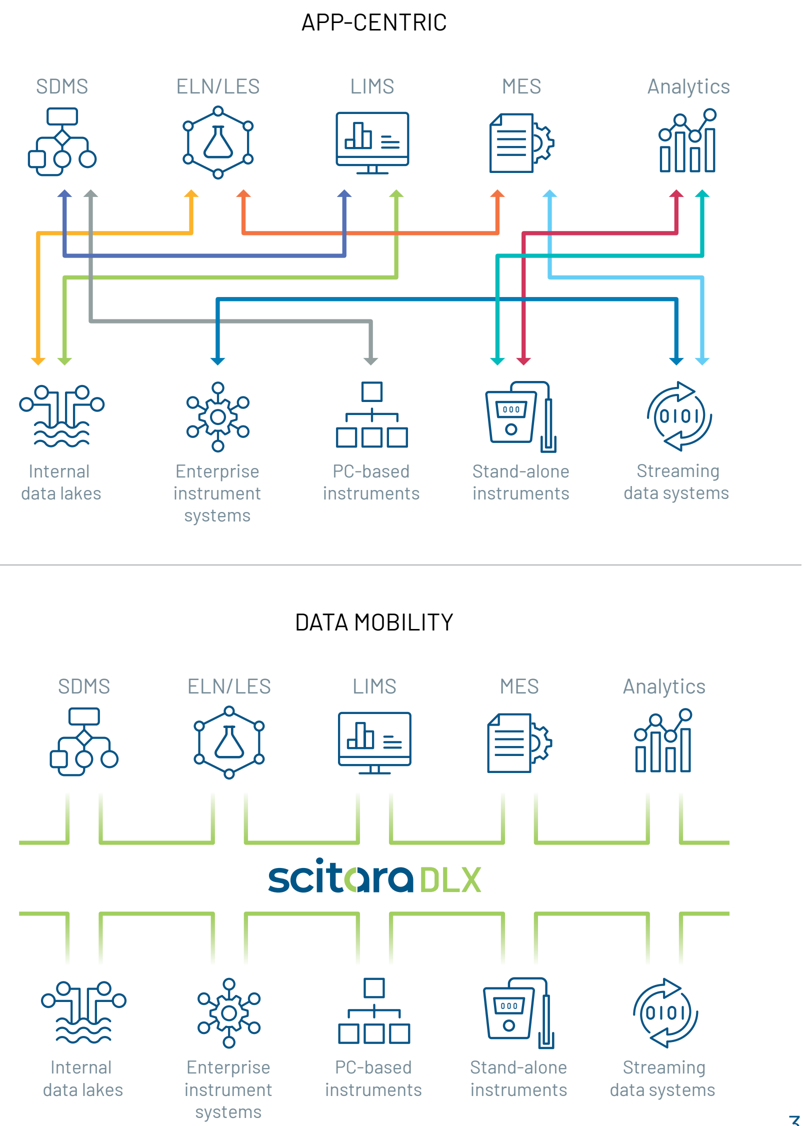
Shifting to data mobility

With the new paradigm of enabling data mobility:

- Data are a key asset of any organization
- Data are self-describing and do not rely on an application for interpretation & meaning
- Data are expressed in open, non-proprietary formats
- Access to and security of the data is a responsibility of the data layer, and not managed by applications
- Applications are allowed to visit the data, perform their magic, and express the results of their process back into the data layer for all to share

The deployment of traditional application-centric architecture has created an enterprise environment where new solutions require complex integrations and many data transformations. Data generated throughout R&D and operations rely on numerous applications and tends to be stored in multiple places. Data transformations are done using many custom parsers or application programming interfaces (APIs) throughout the workflow.

This approach adds unnecessary complexity and fragility to the process. By laying data at the core of an enterprise architecture, IT architects can unlock operational efficiencies and data silos can be knocked down to enable flexible integrations and business agility.



The need for universal connectivity

As previously stated, scientific labs employ a variety of instruments, applications, and services (tools) to fuel innovation and ensure product quality. As the challenges scientists face become more complex, the tools they use to overcome them become more varied and advanced.

Without a universal, enterprise-wide interface for these tools, two fundamental problems arise.

Firstly, data—the lifeblood of science—cannot be transferred easily or efficiently to data consumers or decision makers. Secondly, bi-directional and/or multi-directional data exchange is typically limited to complex, bespoke systems, or vendor-specific silos, in which the vendor controls operations and capabilities within a closed ecosystem.

All these problems significantly slow the pace of innovation in science. In fact, it is estimated that up to 50% of lab time is spent on data preparation and deployment[2]—and it is widely recognized that a solution is urgently needed.

But while other industries have been able to leverage modern technology to improve automation and gain efficiencies, scientific labs face a variety of unique requirements that have made it difficult to follow suit, such as the need for compliance in regulated markets, flexibility in research environments, data integrity demanded by decision-makers and regulators, and a mix of legacy and modern technologies.

Despite these challenges, the dire need for digital transformation within the lab environment has driven innovation that promises to change the game.

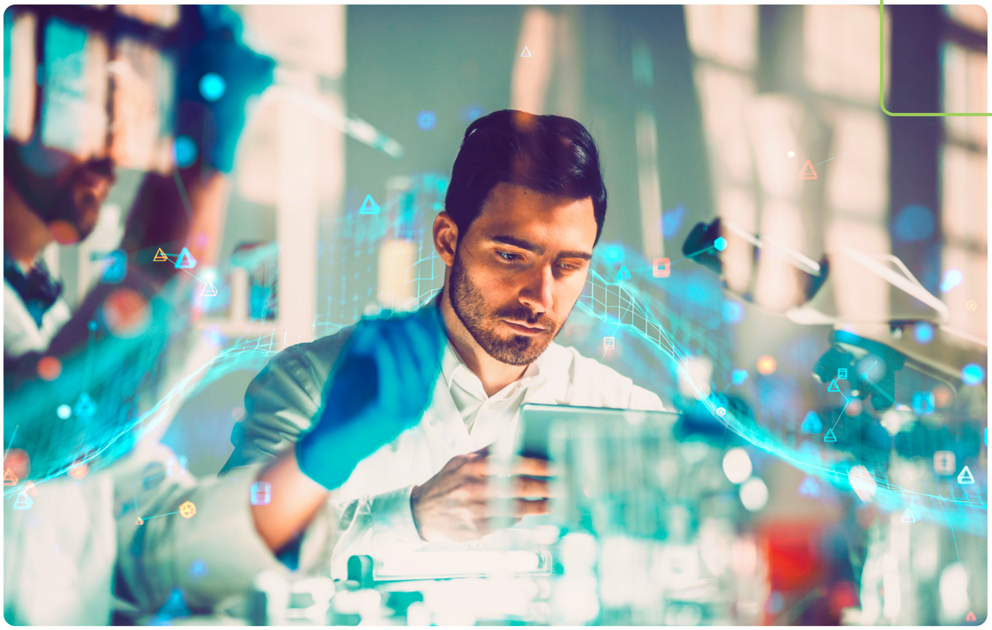
The results are deeply problematic:

- Scientists spend time manually transcribing data between tools, increasing the likelihood of errors.
- Poor connectivity means simple, repetitive workflows cannot be automated, which reduces lab efficiency.
- Tools like AI and ML cannot be fully leveraged because data is not curated or properly formatted, resulting in fewer insights.
- Lab assets are not managed holistically, leading to unnecessary down-time.
- Data archival is bespoke and not automated, which creates data silos and incomplete backups.
- Compliance monitoring is cumbersome and not automated, increasing regulatory risks.
- The Total Cost of Ownership (TCO) of IT-related solutions in a lab environment is too high
- The timelines for implementation and integration of new solutions that require the consumption of data are too long
- Incompatibility with existing technology may limit integration options
- Laboratory operations are inefficient, with many routine or manual tasks
- Scientists don't have on demand access to institutional knowledge
- The greatest company asset, creative people, don't have enough time to innovate

The need for data normalization

Connectivity is the first challenge in the normalization of scientific data for usefulness.

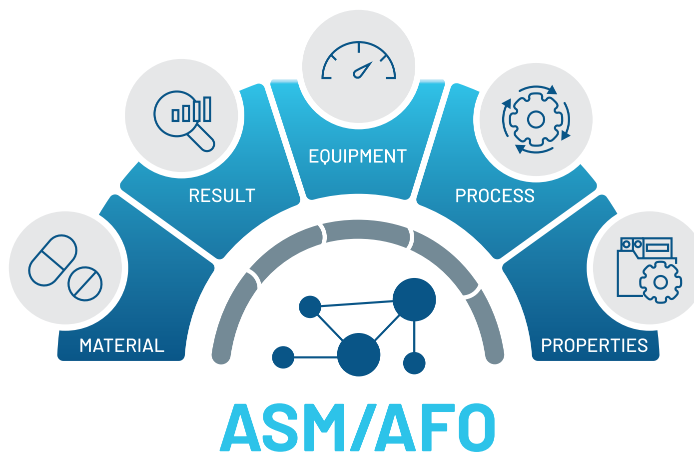
While there are many instruments that may provide the same capabilities and output the same analytical results, the reality is from vendor to vendor the output format can look very different. Even if a scientific laboratory purchased all the same instrument from one vendor, which is not often the case, from version to version, the output data could also be different.



If we take HPLC (High Performance Liquid Chromatography) data as an example, we would expect that a subset of the results would define peaks, retention times, and concentrations. The results from different instruments could return peaks as peaks or peak names, retention times as retention time, time, or R.T, and concentration as concentration or Conc. Each variation of result label naming causes inconsistency in the automated output interpretation. Someone needs to identify and manually map the fields for the data to be used further for analytical purposes.

Data normalization with the Allotrope Simple Model

The Allotrope® Simple Model (ASM) was created by Allotrope Foundation®, a consortium of chemical and life sciences companies, to revolutionize the way we acquire, share, and gain insights from scientific data. The ASM is a JavaScript Object Notation (JSON)-based standard for the structure of instrument data. Through its use of JSON, the de facto standard by which systems on the internet share data, the data in an ASM is designed to be easy to read, write, and transmit by any modern software system. Based upon an intuitive tabular approach, it is a set of key/value pairs, where the keys describe the types of values represented, and are taken from the formal vocabulary defined in the Allotrope Foundation® Ontology (AFO) to ensure consistency and standardization.



With a very rich vocabulary foundation and the largest domain coverage in the industry, the ASM provides an excellent foundation for a data normalization layer. The ASM Models build upon Allotrope's in-depth modeling of over 50 different laboratory domains and make it easy for generators and consumers of scientific data to represent their data in a simple yet comprehensive way that leverages Allotrope's extensive ontologies to ensure a consistent nomenclature across any technique.

The transformative power of an iPaaS for Science:

Scitara DLX is an industry-changing iPaaS solution that connects the world's scientists, instruments, applications, and services.

Scitara DLX:

- Automates the exchange of scientific data across multiple network endpoints through proprietary connector technologies that drive data mobility.
- Creates a flexible ecosystem in which integrations and automations for instruments, applications and services can be switched in and out on-demand as business needs change.
- Accelerates the delivery, accessibility, and exchange of compliant data across the entire scientific enterprise.

  DLX's unique plug-and-play connector technologies ensure unparalleled data connectivity, while flexible, vendor-agnostic configuration tools make it easy to support hundreds of instruments and applications—both old and new.

The language of automated science

With the combination Scitara DLX's ability to provide data mobility and transform data in flight as well as the Allotrope's ASM for the structure of instrument data, result data can be automatically standardized before storage in data lakes without need of further interpretation.

- The DLX engine provides a flexible flow and exchange of scientific data
- ASM/AFO functions as an interoperable data layer (JSON-based standard structure with formal vocabulary)
- DLX can perform proprietary data transformation from a data source endpoint (such as lab instruments) to ASM structure using JSON transformation
- DLX can perform data transformation from ASM structure to a proprietary data destination endpoint (such as ELN) using JSON transformation
- Using ASM as an interoperable data structure enables dataflow reusability
- Only one type of transformation is necessary to be defined in DLX (to/from ASM) once a proprietary data is in a proprietary JSON format
- Scalability and modularity with a library of "ASM/JSON transformation" plugins (connectors)

Scitara author bio

Daniel DeAlmeida

Head of Product Management, Scitara

Scientist and Product Executive with decades of experience connecting technology and business to deliver customer-focused solutions.
BIOVIA- Dassault Systèmes, LabVoice, Spectrum Analytical-Eurofins.

Scitara would like to thank Allotrope Foundation for their critical review of this white paper, especially Amnon Ptashek, Allotrope's Technical Director.

For more information on how to build a connected laboratory:

BOOK A DEMO HERE

scitara
meet the modern lab™

11 Apex Drive, Suite 300A, Marlborough, MA 01752
info@scitara.com P: +1 774-847-5034