New data container construct for automated processing of LC/UV/MS data to support high throughput chemistry

Richard Lee,[1] Andrey Paramonov [1]

[1]Advanced Chemistry Development, Inc. (ACD/Labs), 8 King Street East, Toronto, ON. M5C 1B5. Canada

High Throughput Chemistry and Analytical Data

High throughput (HT) chemistry workflows can often entail complex experimental designs. Depending on the type of experiment, this can range from screening to library synthesis where there may be up to 1536 unique product structures on a single plate.

LC/UV/MS instruments are often the analytical platform of choice for high throughput analysis. However, the analysis step is often the bottleneck for analytical support systems.

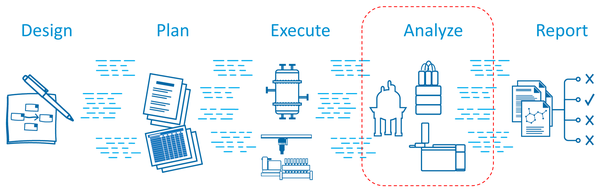

Figure 1. Overview of steps for high throughput chemistry. The major bottleneck is not the chemistry execution step but rather analysis of the chemical reactions.

Here we present a new informatics framework to support the vast amount of high throughput chemistry data acquired on LC/UV/MS systems.

Table 1 proposes an estimated number of low resolution LC/UV/MS data sets acquired in support of high throughput chemistry groups.

Number of Scientists	Experiment Type	Experiments per Month	Analysis Data Sets Per month
150	Batch-based	20	3000 rxns
25	96 well + singletons (1 plate/mth) + singletons	120	3000 rxns
25	96/384/1536 (1 plate/mth)	100	50 000 rxns
			56 000 rxns
	Minimum number of LCUV or LCMS data sets, excluding purification and additional support chromatography		670 000/year

Table 1. Estimated number of HT chemical reactions undertaken by large/medium pharmaceutical organizations, with a minimum number of LC/UV/MS analysis, excluding purification and quantitation.

Based on a data model that supports hierarchal structure for analytical data, the HDF5 format is optimized for low latency and high throughput (input/output). The "chunked" nature of the format allows for quicker access and specific information recall.

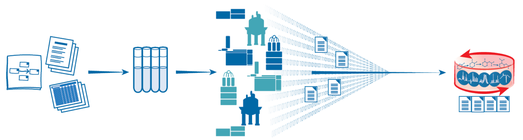

Figure 1. From experimental design to execution, analytical data can be assembled into the new data container where other subsystems can access this data for either query or processing.

High Throughput Chemistry Data Support Framework

In a vast majority of high throughput chemistry experiments, the analysis is typically an LC/UV/MS method that is 3-7 minutes in length. Although this may be short for a single sample, when analyzing a 96 well (or greater) plate the number of samples to be run and analyzed can be significant, often requiring several hours for complete data acquisition.

To mitigate the data processing challenges, data is often processed as soon as the system has completed acquisition. This approach works well for initial data assessment and processing but when reprocessing/rework of the data is required (unanticipated targets, RT shift of target compounds, etc.), the amount of time this takes can be significant due to the multiple steps required to locate, pull, and process the entire set of raw data.

The framework, illustrated in Figure 2, describes the data processing and architecture support for a variety of data servers. This new framework integrates current technologies with a new data container, and a web-based analytical data review interface.

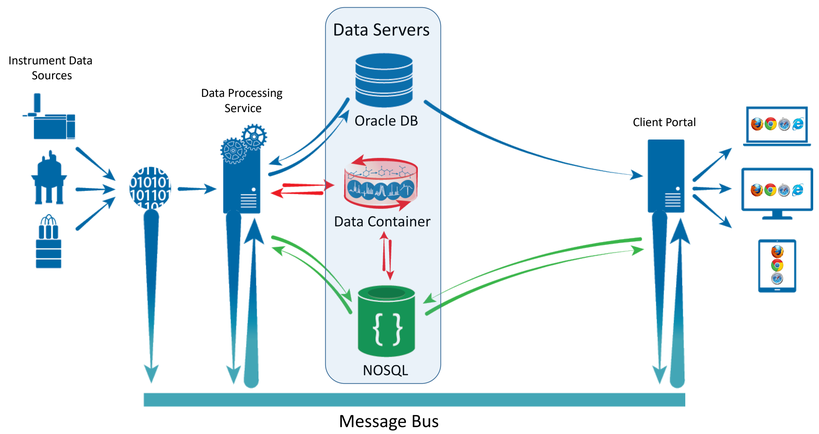

Figure 2. Schematic illustrating the IT architecture of ACD/Labs' new data container as part of an analytical data processing framework.
From left to right:

i. Instrument data sources, either directly from the instrument system or central data location
ii. Data processing service for initiating background activities including: monitoring, data assembly, and data processing
iii. Data container repository for meta data, additional tabular data, analytical data
iv. Client-end web browser for viewing data.

A messaging bus system serves as the back bone for communication between all sub-systems

Web Based Data Review

Katalyst D2D™—a web application for HT experimental design and data visualization—is one example of how the new data container can be leveraged to accelerate analytical data processing. The interface allows the user to lay out various experimental designs. Constructed on a Java script, Katalyst is based on a widget design, allowing for flexible interface configuration, as displayed in figures 3 and 4 below.

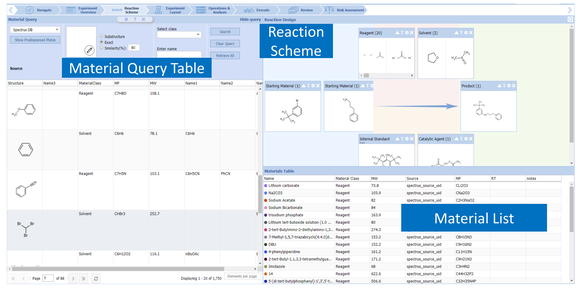

Figure 3. Experimental design space configured in Katalyst D2D. Each window in the graphical user interface is developed as a widget that can be configured. Additional widgets can be added to the layout by the end user.

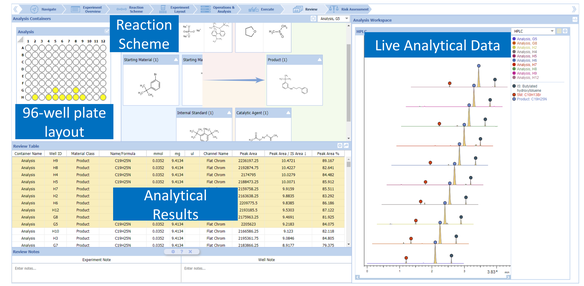

Figure 4. Data Review layout of Katalyst D2D. This layout shows live analytical data connected with each well in the plate. Simply click on the well (or select 'Well ID' from the table)to view the associated chromatogram, mass spectrum, or other analytical spectra.

Preliminary results suggest that this framework, incorporating ACD/Labs' new data containers, has several advantages:

- Access only required information from raw data vs. loading the entire data set
- Decrease data processing time from 40–60 min to sub 5 min, often reaching below 3 min for a 96 well plate
- Based on HDF5 technology, affords scalable storage and deployment
- Can store complete experimental design data (chemical structures, meta data), raw analytical data, processed data, and numerical results in a single container
- Along with LC/UV/MS data, the new data container can also store additional analytical data such as NMR, UV, XRPD, etc.