Data: Do We Have Too Much of a Good Thing?

Sanji Bhal, ACD/Labs

Can there be too much of a good thing? When it comes to data, the answer may be yes. It seems the more data we have, the more knowledge we lose; the more value goes unrecognised. Over the past decades we've gone from wet lab methods, to high-powered spectroscopy, to lab-on-a-chip, and more. The chromatography community knows this all too well. As the science of chromatography has advanced (for example from HPLC to UHPLC) so too have the numerous instruments and software become more sophisticated. The data generated from new tools and techniques have been massive and have proved invaluable to the advancement of science, without a doubt. In fact, organisations' investment in instruments now exceeds $1 million annually, and 80% of organisations rely on analytical data for decision-making.

But as we've all experienced, the massive amounts of data we are generating can be a real pain. Or, more precisely, managing the data can be a real pain. It comes in different formats from different instrument vendors; it's shared differently and is more and more often spread over multiple departments, geographic locations, and between partner organisations/contractors. If the data generated from method development and optimisation could be re-used effectively, imagine the reduction of workload and increased productivity from your lab. Of course, this isn't a problem relegated to chromatographers or analytical chemists in general. R&D labs across disciplines are grappling with the glut of information and how to best extract its value.

Is the problem of data overload, and more importantly data management, all that bad, though? Are we really just collectively whining? After all, with the preponderance of data over the years, there have been a number of software management systems and techniques introduced to solve the problem.

At ACD/Labs we set out to gather some empirical evidence on data management. In 2015 we approached scientists, managers, and executives in R&D organisations to complete a survey that would provide a clearer understanding of the knowledge of analytical data management. What we found was unsurprising given what we'd heard from customers at the outset of the project, but at the same time the results were a little incredulous.

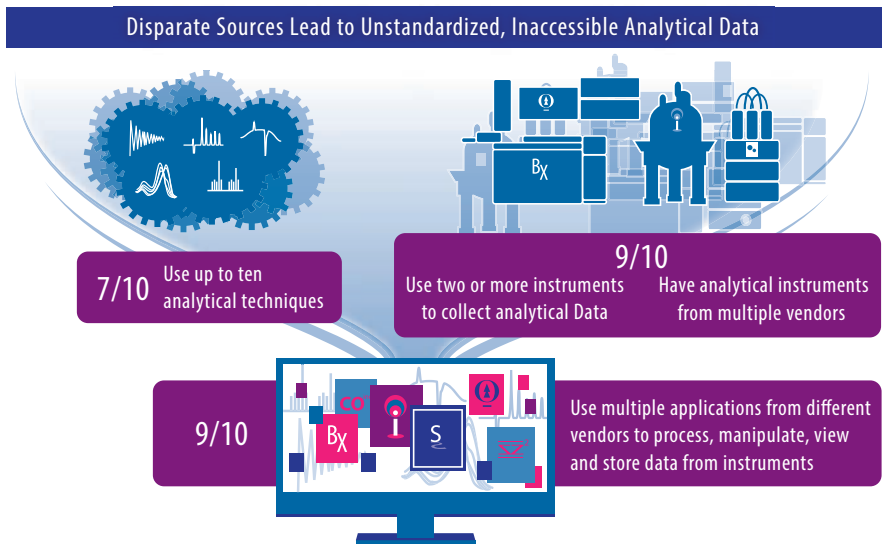
It turns out that an overwhelming 70% of survey respondents use up to 10 different analytical techniques, as well as a variety of instruments from multiple vendors, to collect and analyse their data. And a majority of respondents indicated that once information is extracted after the analysis, there is no one standard method used to share or record the data.

There's more. Remember that old childhood game where a kid whispers a story in another kid's ear, and that story is passed along in a circle until the last child in the circle has to repeat it out loud? Then everyone laughs because the story is nothing like it was when they heard it, some pieces of information are missing and others are incorrect.

That's happening in R&D labs, too. More than half, 55% to be exact, of survey respondents indicated they relied on conversations to share data. Paper documents and images ranked high as tried-and-true methods as well.

Not only is data transferred by paper and conversation, but it's stored in multiple, disconnected places. Our survey found that nine out of ten companies rely on different brands of equipment for analysis, each with their own systems that are incompatible with those supplied by other vendors. In addition to the problem of having 'islands' of information strewn across an organisation, staff must be trained on multiple systems, not just when the systems are implemented but when they are updated throughout their lifecycle.

To be sure, what's occurring in R&D organisations is not nearly as laughable as the storyline changes in a simple game of Chinese whispers. Analytical data is relied upon to identify and characterise samples, and to address what can become very significant business problems, such as patent protection, inquiries from regulatory bodies, and impurity resolution. Think about the numerous different ways separations are used in R&D for both purification and analysis, from preparative HPLC, to quality control



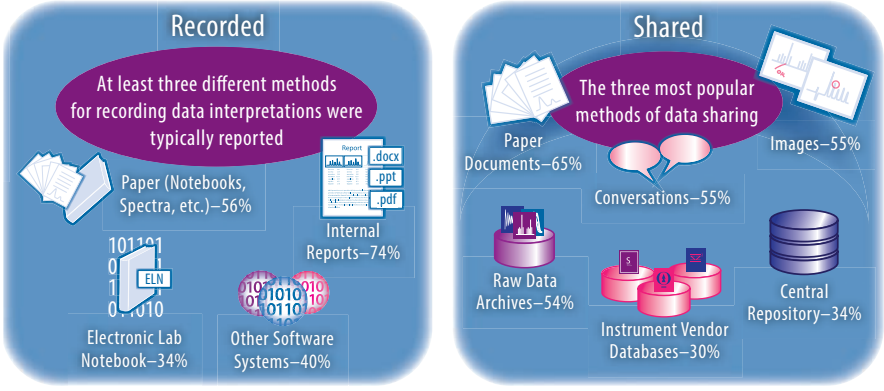
in manufacturing. Putting the problem into another perspective, we can think about the practical ramifications of an FDA question from a new drug application (NDA). When data cannot be accessed or identified quickly to answer the question, scientists may be forced to re-synthesise compounds to start their analysis from scratch—at just the same time in the drug development cycle when every day costs the organisation money.

Making matters worse, more than half of chromatographic work and analytical data analysis today is outsourced. When data is returned to the parent company it's either 'dead data', in the form of an image with accompanying numerical and text data, or in some cases raw data that is incompatible with the parent company's systems. The result is that it becomes nearly impossible to explore the data and analyse it for anomalies, or to answer new questions.

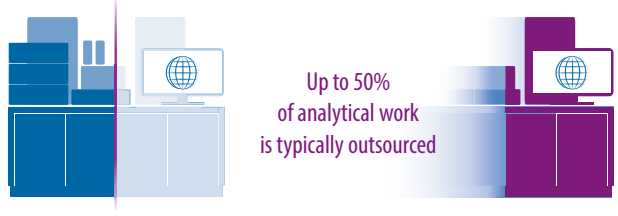
The survey revealed that the importance of analytical data is undisputed. 80% of respondents agreed that they rely on analytical data for decision making and that their organisation would benefit from better searching and sharing of analytical

Spectral Data Interpretations are Not Recorded or Shared Uniformly

Information Extracted from Analytical Data is Scattered, Unsearchable, and Difficult to Re-Use



Problems Are Exacerbated by the Trend Towards Increased Outsourcing



data. With 60% responding that their organisation should invest in additional data management technology it seems clear that many of the existing technologies have failed to live up to their promise to manage the deluge of disconnected, unsearchable, and difficult-to-re-use data. It is unrealistic to assume that any one technological solution can manage all the different kinds of data generated in R&D organisations. Rather, specialised systems that integrate and complement each other are the answer. With analytical data being of critical importance, organisations should invest in technologies that speak the specialised language of chromatographic, spectrometric, and spectroscopic data.

One of our customers, Steve Thomas, investigator at GlaxoSmithKline, summed it up this way: "We are using analytical information like a jigsaw. Getting the information in one place is vital, and that used to be in people's heads. But when a colleague retired, that information was lost. We needed a repository of information that didn't forget, didn't go senile, and didn't go to a competitor."

Thomas has implemented ACD/Labs' data management platform to enable his teams to access and manipulate data from any source. The system's ability to keep records and retain information meant the team was able to extract the knowledge when they needed it. "All the data is there, filling its place in the jigsaw," he said.

From our perspective, what's needed is a vendor agnostic, unified approach to data management, one that allows data to be managed in a centralised place and reviewed and re-interrogated when necessary. Only when we are able to easily access, understand, and share data will organisations' $1 million annual investments into generating data truly return their full value.

Analytical Data is Indispensable

Analytical data is routinely relied upon to identify and characterize samples and address low frequency, high impact problems such as patent protection, impurity resolution, and responses to questions from regulatory bodies.

8/10
We rely on analytical data for decision making

8/10
My organization would benefit from better searching and sharing of analytical data

7/10
Sharing data and interpretations between departments/ partners is valuable to my organization

6/10
My organization should invest in additional analytical data management technology