# The State of Analytical Data Management

Sanji Bhal, ACD/Labs

Over the past decade, we've experienced immense change in R&D. From new instruments and software to the rise of the digital lab, the landscape of analytical chemistry data has evolved. At ACD/Labs, we need to keep a finger on the pulse of analytical data management (ADM).

That's why we launch a comprehensive survey every few years to uncover the latest trends and preferences regarding analytical chemistry data and its management.

In this survey we heard from academia (30%), industry/manufacturing (26%), biotech/pharma (13%), government (11%), nonprofit research (9%), contract service providers (5%), and consultants (3%).

Here's what we found…
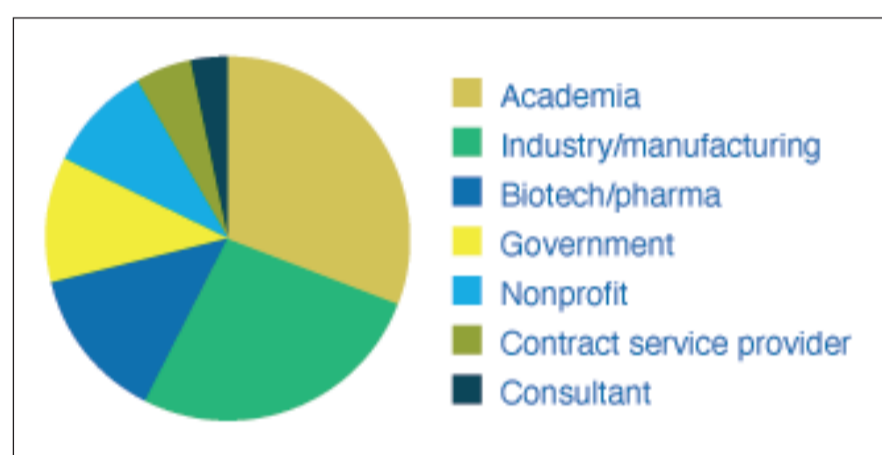

Data Diversity is a Real Problem and a Necessity

Data is the backbone of all scientific research projects. Analytical data is primarily collected to ensure the identity and quality of materials to specifically:

1. Understand the structure or composition of materials and processes by which they're made

2. Evaluate the performance of experiments, materials, or processes

It is often necessary to run several different analytical experiments to answer these questions (e.g. LC/MS and NMR). Analytical labs are equipped with a variety of instruments so that analysts can choose the best instrument for the answers sought. Since we want to execute the best science, diversity of analytical data is necessary. In addition, many research teams use instruments made by multiple vendors, which leads to file compatibility issues.

Unsurprisingly, our survey found that over 92% of respondents collect data on numerous instruments, use multiple techniques, and rely on diverse software for processing analytical data. To break this down further, 45% typically use 2-4 analytical techniques; 37% use 2-4 different instruments to collect data (33% use 5-9 instruments!); and 54% use 2-4 software applications to process their data.
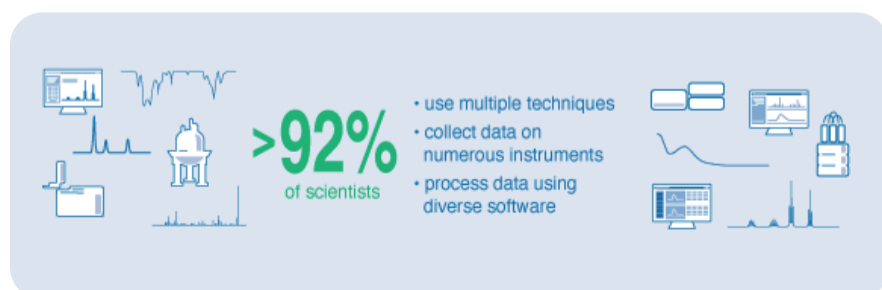
Analytical Data is Managed in Multiple Applications and Shared Haphazardly

The diversity of analytical data means that it's stored and managed in many different applications and systems for most organisations.
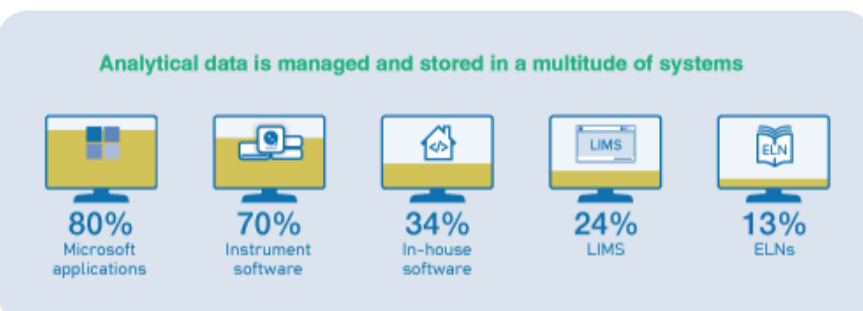
Microsoft applications are still the most popular way to manage and share analytical results, selected by 80% of respondents. Whether its Excel spreadsheets, PowerPoint presentations, or email, ubiquitous access to these applications makes them an easy choice even though they are neither designed nor best-suited for scientific data sharing and management.

Instrument software was the second most popular choice at 70%. While instrument software is restrictive to only processing and analysis of the data collected on that instrument, it is designed for it. It was surprising to learn that many organisations are still using software developed internally to manage and share analytical data, even with the development and maintenance overhead required. Many other systems deployed in R&D are also used to house and share analytical data ELNs, LIMS, CDSs, SDMSs, archives and more.

These systems represent different activities and are often used in combination throughout the lifecycle of an analytical data file:

1. Stored in a raw data archive to affirm quality and accuracy

2. Processed and stored in vendor software to extract results and retain the processed data file

3. Results shared with scientists via LIMS or Email may include images of spectra, confirmation of expected structure/material composition, and text results (MW, peak tables, retention times, etc.).

4. Decisions based on those results may be recorded in a scientist's ELN along with the image of a spectrum, peak table(s), confirmation of expected structure with the scientist's notes. Decisions may also be presented in an internal meeting via PowerPoint and subsequently stored on SharePoint, or shared in a report

5. Stored in the CDS or SDMS to conform to FAIR/ALCOA principles and fulfill regulatory requirements

Table 1. High-level advantages and deficiencies of applications and systems typically involved in the analytical data lifecycle

Application/System	Advantages	Deficiencies
Instrument vendor software	• Store raw and processed data • Available for (re-)processing and posterity	• Lacks experimental context • Requires access to specific versions of software
Microsoft Applications (PowerPoint, SharePoint, Teams, Outlook, Excel etc.)	• Widely available	• Scientific context is not supported • Data abstracted to text & images
In-house developed software	• Fit-for-purpose	• Requires dedicated staff for maintenance
LIMS	• Sample-centric storage	• Lacks experimental context • Data abstracted to text & images
ELN	• Experiment-centric storage	• Inability to review and reprocess data • Data abstracted to text & images
CDS	• Assemble multi-vendor data • Data is "locked down" for regulatory-readiness	• Limited ability to review and reprocess analytical data • Data lacks experimental context • Data abstracted to text & images
SDMS	• Assemble multi-vendor data • Data is "locked down" for regulatory-readiness	• Limited ability to review and reprocess analytical data • Data lacks experimental context • Data abstracted to text & images

Scattered data makes assembly, a crucial step in decision-making, difficult as it forces scientists to search multiple locations for answers. When there are many possible locations for data, the path of least resistance is often to repeat the experiment or request the data from a colleague, which wastes time, materials, and can cause frustration.

## Scattered Data Makes Reporting Time-Consuming

Reports are a key way to share information within an organisation or with external partners. Only 18% of respondents said they rarely (or ever) collate analytical reports with data from different instruments and techniques, and 40% do so weekly or daily. How much time is wasted, then, in moving from system to system to collect all of

40% collate analytical data from different instruments and techniques daily/weekly
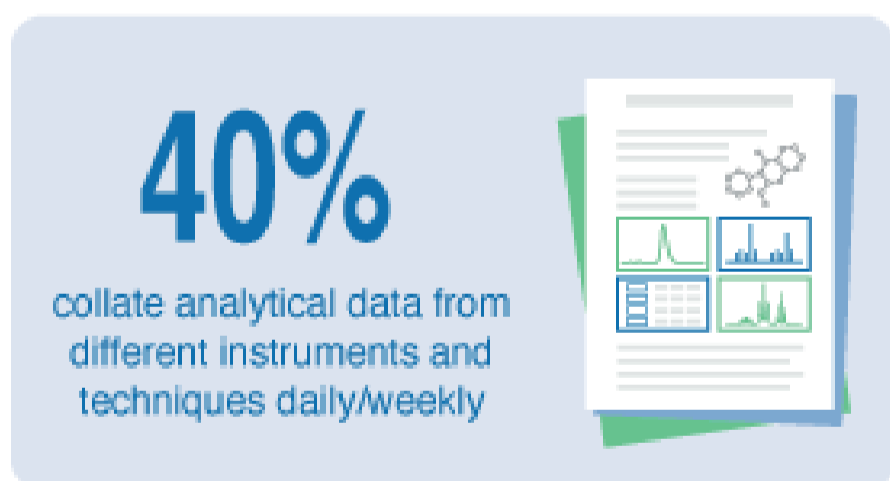
the relevant data to compile these reports? By fully implementing an ADM solution, scientists can collate reports by simply linking to data.

Analytical Data is Mission-Critical but Difficult to Access and Share

Nine out of ten respondents note that they need NMR, LC/MS, GC/MS, or other analytical data daily to make decisions. Seven out of ten agree that sharing data and interpretations within their organisation is important. However, for an element that is mission-critical to their job, it's not easy to access or share that data with others; 50% agree that searching for data in their organisation is a challenge, while 68% say it's hard to access and share with others.

9/10 need analytical data on a daily basis

68% say data is hard to access and share
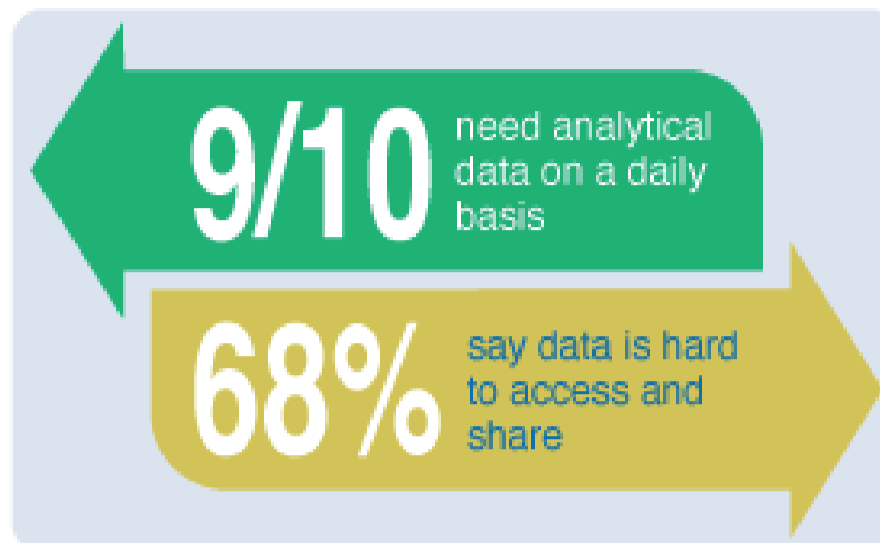
Data access is especially challenging when it involves data collected by others, in larger/scattered organisations, or when a team member leaves or joins the organisation.

Reasons for Needing Access to Data from Past Experiments

To appropriately address these hurdles to data access, it's essential to consider why scientists need to access data from past experiments.
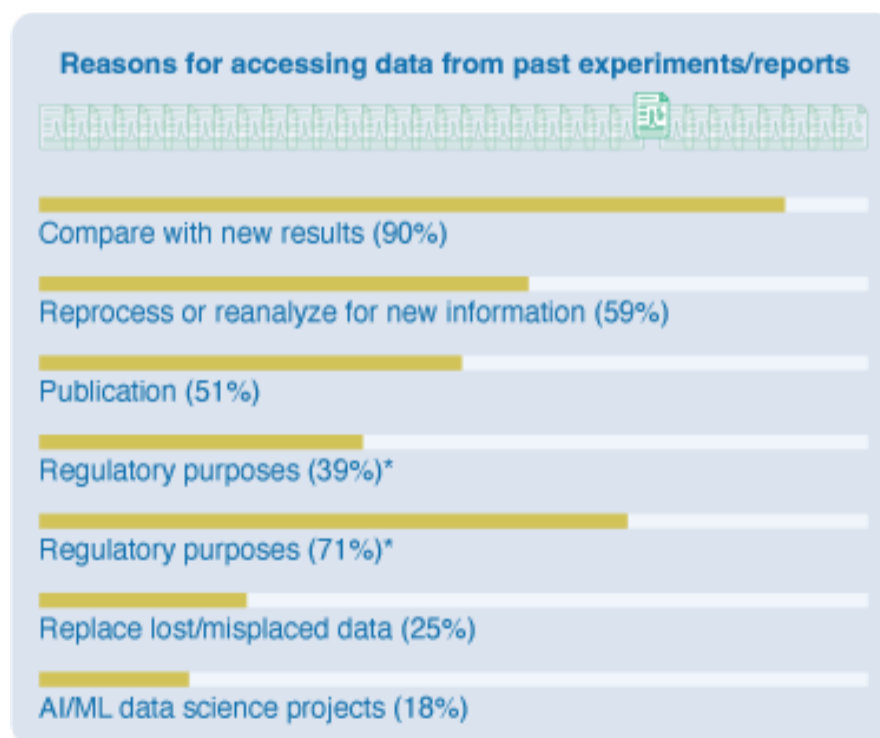
Across all R&D sectors, the top three reasons for accessing data from past experiments and reports are:

1. To compare with new results
2. To reprocess or reanalyse for new information
3. For publication purposes

And, the top 3 reasons for Pharma/BioPharma are:

1. To compare with new results
2. For regulatory purposes
3. To reprocess or reanalyse for new information

Accessing past data for regulatory purposes was the second most important factor for Pharma/Biopharma, while it was understandably insignificant for academia and non-profit organisations. Other than this variation, the reason why data from past experiments needs to be accessed was consistent across the R&D sectors.

Reasons for accessing data from past experiments/reports

- Compare with new results (90%)
- Reprocess or reanalyze for new information (59%)
- Publication (51%)
- Regulatory purposes (39%)*
- Regulatory purposes (71%)*
- Replace lost/misplaced data (25%)
- AI/ML data science projects (18%)

25% of respondents accessed old data to replace data that was lost or misplaced. Properly managed, accessible data can deliver significant savings of time and effort. Especially when old data cannot be found, the alternative is to re-run experiments!

Only 18% of respondents accessed old data for data science projects.

While academic and non-profit groups may focus less on data management than other R&D sectors, this could be an opportunity to improve productivity. From my own days in the lab several research projects would pass from one student to another. Finding data from within the group, even from current colleagues, was challenging.

Opportunities for Improvement

Cloud-based data management is increasingly enticing to streamline storage and access

~50% feel that cloud-based data management is important
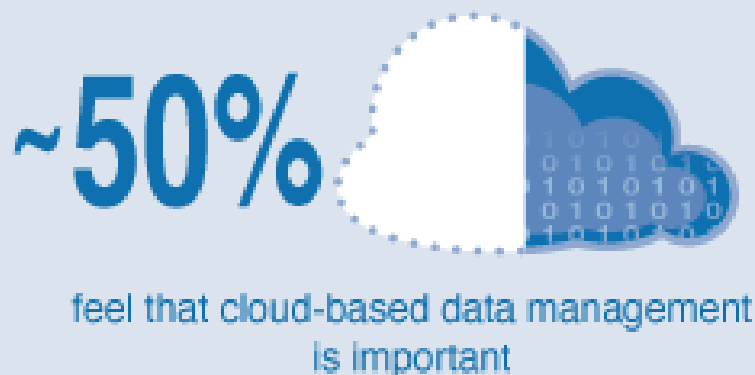
Scientific R&D is on the edge of a cloud revolution. Cloud-based data management offers streamlined collaboration with information available to everyone, regardless of location. In addition to reducing IT maintenance overhead, cloud-based storage provides fast scalability, and added data security. In the long run, more immediate access to data means increased ROI and reduced expenditure.

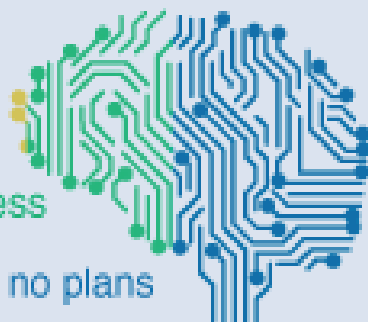Nearly half of respondents (47%) also agreed that cloud-based data management solutions are important.

Advanced Technologies - Like AI and ML - Are Appealing but Few Have Implemented Them for Analytical Data

Use of Analytical Data in AI/ML

6% fully implemented
43% in process
51% no plans

There has been a lot of hype about artificial intelligence (AI) and machine learning (ML) over the last few years. While there is plenty of potential for these technologies in the life sciences, our results reveal that the industry is years away from full implementation.

Only 6% of respondents' organisations had fully implemented the use of analytical data for data science projects, while 43% are in the process of doing so. 51% have no plans of using analytical data for AI and ML projects.

There is a lot of disparity around the implementation of AI and ML. This is unsurprising, especially given the daily volume of analytical data and its diversity. The cornerstone of data science projects is curated, normalised data, which is challenging for analytical data. If taking advantage of AI is a long-term goal, it's important to identify how analytical data fits into that goal and start at the beginning. Many analytical data management solutions in place today do not prepare that data for use in data science projects. Automation to gather data without burdening scientists and internal agreement on how that data will be normalised, are critical first steps.

There is Work To be Done…

70% of respondents agree their organisation needs investment in newer/better data management technologies.

70% want more investment in data management technologies

Specific improvements mentioned, include:

- A centralised system to manage data versus several programs or systems
- User-friendly systems
- Cloud storage compatible
- Stronger data security

For almost three decades ACD/Labs has been helping R&D organisations manage their analytical data and improve the efficiency of workflows where analytical data is key. More recently, supporting the migration of software and solution deployments to the cloud and consulting on data preparation for use in data science (AI and ML) projects.

Want to see how your organisation stacks up, or learn what ACD/Labs' solutions can do for you? Request a free consultation at acdlabs.com or meet with us at Pittcon (Booth 2355).

Read, Share and Comment on this Article, visit: www.labmate-online.com/article