JOINING UP THE LABORATORY

Peter Boogaard reviews efforts to make the laboratory an integrated operation

Nixx Photography/Shutterstock

It is easier to get data into scientific databases than to get valuable information out of it. For years, we have been spending time and money to integrate systems and processes in the laboratory's knowledge value chain. Many laboratory integration projects are under pressure to deliver on their expectations, as defined at the kick-off of. So why is it that laboratory integration is so difficult? What are the obstacles to creating value for the consumers of the laboratory data? Do we know what these users need and how they would like to consume this information?

Imagine that in the music world, each label has its own proprietary music file format. How would you be able to share music? By default, standards make it easier to create, share, and integrate data. Do we know the requirements of such a data standard? What about managing metadata-controlled vocabularies? Data standards are the rules by which data are described and recorded. In order to share, exchange, and understand data, we must standardise the format (data container) as well as the meaning (metadata/context). As of today, there is no unified scientific data standard in place to support heterogeneous and multi-discipline analytical technologies. There have been several attempts but they are limited in scope, not extensible or incomplete, resulting in recurring, cumbersome and expensive software customisations.

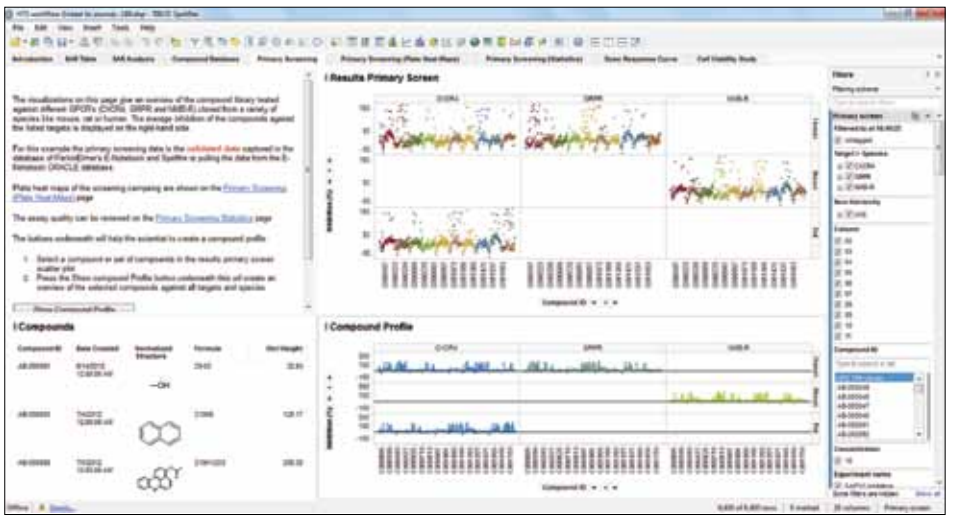
PAY ATTENTION TO THE CONSUMER OF THE DATA

Integrating laboratory instruments started when instrument vendors, such as Perkin-Elmer and Beckmann Instruments, created the first laboratory information management system (LIMS) software, in the early 1980s. The initial objective was to support the laboratory manager with tools to create simple reporting capabilities to enable the creation of simple certificate of analysis (CoA) reports. These systems were initially designed to support a single consumer, namely the scientists and lab managers. In today's world, consumers of laboratory data can be found across the entire product lifecycle, and may include external organisations such as CROs and CMOs (Table 1). A different mind-set is required to adapt to this expanded view of the world. It is critical to first analyse who these new lab-data consumers are, and get an understanding of what their objectives are. Often forgotten, but as important, is to investigate what their perspective is on usability. The newcomers may be a non-technical audience! Stephen Covey phrased it very nicely: 'Seek first to understand…. And then to be understood.'[1] It may sound obvious, but it still remains a valuable statement before starting any automation project.

Stephen Covey phrased it very nicely: 'Seek first to understand…. And then to be understood.' It may sound obvious, but it still remains a valuable statement before starting any automation project

Table 1: Selected consumers of laboratory information data

Consumer	Objective	Impact / benefit
Patient	Assure secure instant access to medical data for doctors.	Better healthcare at lower cost
Fellow scientist	Re-use experimental data and leverage learning. Higher efficiency and quality. Consistent meta and context data	Higher efficiency and quality
Legal	Protect company IP	Consistent externalisation processes (CROs)
Finance	Understand overall life-cycle cost of operation	Holistic overall view
Customer care	Product complaints and product investigations	Secure branding image of company
Regulation	Faster responses to compliance inquiries	Simpler mechanism to audit heterogeneous scientific data
Management	Identify areas for continuous improvement in process. Reduce costs	Risk-based information across heterogeneous data systems
Stability labs	Simpler mechanism to create e-submissions. Ability to submit standardised e-stability data packages	Faster responses during studies, Increased efficiency
CRO/CMO	Focus on lowering cost/analysis by decreasing IT complexity and overhead	Acceleration move from paper to 'paper-on-glass'
IT	Reduce bespoke/custom systems. Consolidation of systems. Reduce costs	Unified systems. Simplify IT processes



Example of dashboard

- For the **scientific researcher**, the ability to record data, make observations, describe procedures, include images, drawings and diagrams and collaborate with others to find new chemical compounds, biological structures, without any limitation, requires a flexible user interface.
- For the **QA/QC** analyst or operator, the requirements for an integrated laboratory are quite different. A simple, natural-language based platform to ensure that proper procedures are followed will be liked.
- To investigate a client's complaint professionally, the **customer care** employee requires a quick and complete dashboard report to look at metrics for all cases, assignments, and progress in real-time, by task, severity, event cause, and root cause. The devil is in the detail, and that's where the laboratory data may give significant insights.
- **Legal:** Instead of saying 'we saw that a couple of years ago, but we don't remember much about it', sensitive information can searched and retrieved, including archives.
- During **regulatory inspections** 'show me all the data during this time frame, which raw material batches were involved and show me all the details'.

HETEROGENEOUS SCIENTIFIC CHALLENGES

The lack of data standards is a serious concern in the scientific community. It may seem a boring topic these days, but the need for standardisation in our industry, has never been higher. Without such standards, automating data capture from instruments or data systems can be challenging and is expensive. Initiatives such as the Allotrope Foundation[2] are working hard to address these badly needed common standards.

The Allotrope Foundation is an international not-for-profit association of biotech and pharmaceutical companies building a common laboratory information framework for an interoperable means of generating, storing, retrieving, transmitting, analysing and archiving laboratory data, and higher-level business objects such as study reports and regulatory submission files. The deliverables from the foundation, sponsored by industry leaders such as Pfizer, Abbott, Amgen, Baxter, BI, BMS, Merck, GSK, Genentech, Roche and others, are an extensible framework that defines a common standard for data representation to facilitate data-processing, data-exchange, and verification. One of the ultimate goals is to eliminate widespread inefficiencies in laboratory data management, archival, transmittal, and retrieval, and to support a start-to-finish product quality lifecycle, which would enable cross-functional collaboration between research, development, quality assurance and manufacturing.

The framework will include metadata dictionaries, data standards, and class libraries for managing analytical data throughout its lifespan

The framework will include metadata dictionaries, data standards, and class libraries for managing analytical data throughout its lifespan. Existing or emerging standards will be evaluated and used as appropriate, to avoid 'reinventing the wheel'. It is a wake-up call for the industry but one that may be muted by our risk-avoiding, sceptical mindset. It is reminiscent of how the database technologies emerged in the 1970s. The reader is challenged to identify the similarities between the development of SQL, and the initiative to create an intelligent and automated analytical laboratory.

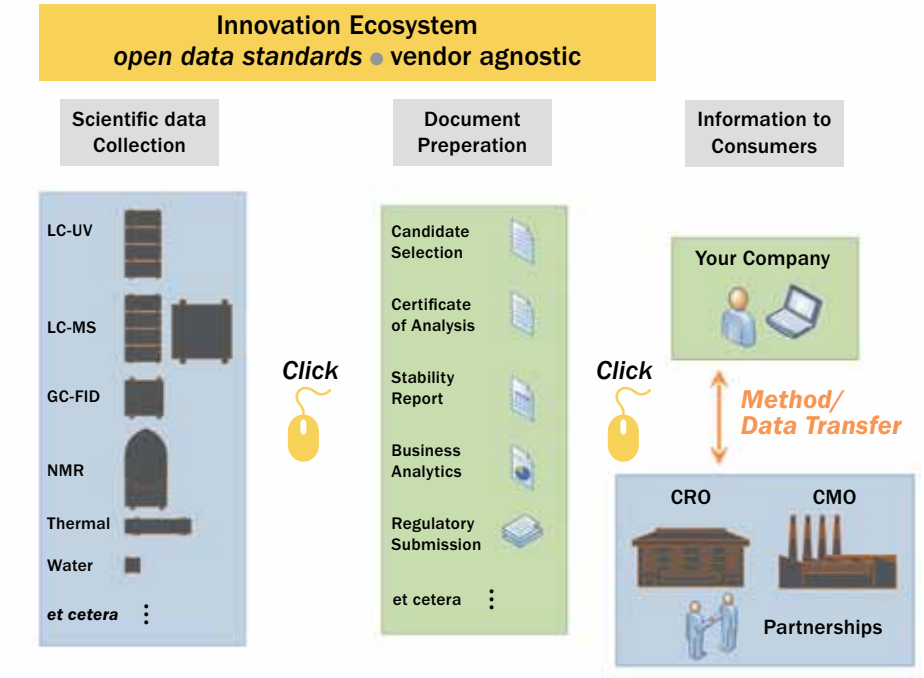
GLASS HALF FULL OR HALF EMPTY?

The deployment of computerised database systems started in the 1960s, when the use of corporate computers became mainstream. There were two popular database models in this decade: a network model called CODASYL; and a hierarchical model called IMS. In 1970, Ted Codd (IBM) published an important paper to propose the use of a relational database model. His ideas changed the way people thought about databases. In his model, the database's schema, or logical organisation, is disconnected from physical information storage, and this became the standard principle for database systems. Several query language were developed, however the structured query language, or SQL, became the standard query language in the 1980s and was embraced by the

Table 2: SQL pros and cons

Why traditional hierarchical was initially abandoned	The SQL a success story
Complex architecture	Extensible open architecture
Slow responses	Split physical & meta data
Vendor bound	Product independent
Inflexible and fixed data schemas	User definable flexible ad-hoc queries capabilities
Required mindset change	Availability of faster computers and networks
Invasive technology	Single database language

Allotrope Foundation



The benefits of open data standards

entire industry. Vendor specific proprietary extensions (e.g. PL/SQL[3]) were allowed in the concept, allowing individual vendors to extend capabilities.

Now back to the laboratory. The current situation is that there is no framework for scientific data standards. Formats are vendor bound, product dependent, and in many cases based upon a closed architecture and are complex in nature. There are plausible reasons why, at this moment, our industry has no general accepted raw-data and metadata standards, but should we not learn from other industries and adopt best practices?

The Analytical Information Markup Language (AnIML) is the emerging ASTM XML standard for analytical chemistry data. The project is a collaborative effort between many groups and individuals and is sanctioned by the ASTM subcommittee E13.15.[4] An AnIML is a standardised data format that allows for storing and sharing of experimental data. It is suitable for a wide range of analytical measurement techniques. AnIML documents can capture laboratory workflows and results, no matter which instruments or measurement techniques were used.

E-Workbook Suite (IDBS) allows spectra files to be dropped in from the experiment whereby they are automatically converted to AnIML and rendered. The rendering application then allows the scientist to annotate the spectra with searchable chemical structures, text, hyperlinks to other systems and records. The AnIML data is also indexed alongside everything else allowing specific searching of metadata and properties. These processes are non-invasive meaning that the originals raw data files are also kept.

An application programming interface (API) specifies how some software components should interact with each other, allowing customers and third parties to extend the types of spectra that are supported by writing new raw data to AnIML converters or plug-in in third party components.

Other examples of changes in the way laboratories may operate in the future relate to how balance and titration instrument vendors are increasing the value of their instruments by implementing approved and pre-validated methods in their instruments. This may sound a small step, but it may have a significant impact on validation efforts in the laboratory and manufacturing operations, such as fewer points of failure during operation, less customisation of software and better documentation.

The desire to convert manufacturing processes from traditional batch-oriented processes to a continuous operation has accelerated process analytical techniques (PAT) technologies as a way to create sustainable and flexible approaches for manufacturing operations. PAT is expected to grow significantly in the next decade. Over time, in-line, @line and on-line analysis will complement and potentially substitute off-line (batch oriented) laboratory manufacturing processes. International regulatory authorities such as ICH, FDA and ISPE are evaluating these new processes intensively and developing new workflows. These processes will have a high impact on how QA/QC laboratories will operate in next decade. International industry standards such as ANSI/ISA-88 (covering batch process control) and ANSI/ISA-95 (covering automated interfaces between enterprise and control systems), are commonly used in manufacturing. By

E-Workbook Suite (IDBS) allows spectra files to be dropped in from the experiment whereby they are automatically converted to AnIML and rendered

Table 3: Applying standards requires a different mindset

Glass half empty	Glass half full
The market is too disperse	Technologies are emerging rapidly
Technology not available	XML and AnIML are accepted as standards
Vendor protection	Empowered customers
Poor performance	Consistent unified long time archive process

References

[1]The 7 Habits of Highly Effective People – Stephen Covey 1990

[2]Allotrope Foundation – www.allotrope.org

[3]PL/SQL (Procedural Language/Structured Query Language) is Oracle Corporation's procedural language extension for SQL and the Oracle relational database

[4]Subcommittee E13.15 on Analytical Data - www.astm.org/COMMIT/SUBCOMMIT/E1315.htm

[5]Technology Management In The Age Of The Customer - Forrester research 2013

Table 4: Potential integrated laboratory killer apps

Laboratory integration killer apps	
Archiving	The perception of data archiving is often related only to storing data. Having meta-data standards as part of the archive procedure, will enable data to be re-used for collaboration between different instruments in-house or externally with CROs
Data finder and data viewer	The ability to do full context searching across heterogeneous data sources, across in-house and external data systems and archives and display in unified viewer
Regulatory reviews	The ability to build and transmit to regulatory agencies a standard data package for inspection without altering the underlying information (e.g. regulatory submissions, stability studies)
Reduction in (re)-qualification processes	In a GxP environment, the ability to automatically update USP methods across individual instruments will significantly reduce the requalification process

incorporating these standards, scientists will be able to mine information from development and manufacturing for improved process and product design. In addition, information is more readily transferable between systems. For example, a recipe delivered in early development can be rapidly transferred to a lab execution system for API manufacture and then to a method execution system for mainstream manufacturing. ERP and MES applications are using these standards and it is very likely that integrated laboratory data management capabilities will be included within their software capabilities.

CONCLUSION

Empowered customers are disrupting every industry. Technology managers must broaden their agenda to consider not just infrastructure and traditional internal IT processes, but also activities to ensure they deliver value for their 'client'.[5] The power of an integrated laboratory environment is its ability to find detailed answers to support the overall business process. It is pure waste to perform labour-intensive hunting for information across multi-vendor, multi-technique databases, manual transcription checking and to manually create reports. Having a common industry standard framework will decrease process variability resulting in better quality and overall consistency. Non-invasive processes have proven to be successful in other industries. It is now up to the industry, regulatory bodies and vendors of scientific instrumentation and software platforms to make it happen. Integrating laboratory information really means integrating scientific data collected in the laboratory and beyond. Time will tell if this industry has been able to adopt such a strategy.

Peter Boogaard is an independent laboratory informatics consultant