Building a Smart Laboratory 2018

An introduction to the integrated lab

From the publishers of

SCIENTIFIC COMPUTING WORLD

www.scientific-computing.com/BASL2018

Pharma&Biotech

Lonza

The MODA™ Platform

The Missing Piece in Your Lab Systems Portfolio

MODA-EM™ Software for QC Micro – Implement, Validate, Integrate Seamlessly.

Lonza's MODA-EM™ Software is purpose-built for QC Microbiology's unique challenges. MODA™ Software delivers better business alignment with QC, at a lower total cost of ownership versus customizing your existing LIMS or other lab systems.

- Reduce investigation and reporting time with rapid analysis and visualization of Micro data
- Manage, schedule and track large EM and utility sample volumes
- Reduce errors while improving compliance and data integrity
- Deliver measurable return on investment from time, cost and compliance savings

Find the missing piece in your lab systems portfolio.
Visit **www.lonza.com/moda** or contact us at **moda@lonza.com**

www.lonza.com/moda

Contents

WELCOME TO THE SMART LABORATORY

This year's edition of *Building a Smart Laboratory* discusses the importance of developing a robust strategy for the deployment of paperless lab technology. As the article on page 6 discusses, in order to gain the most insight and value from paperless technology there needs to be a consistent and comprehensive approach that covers the four most important pillars; connect, manage, decide, archive.

As laboratories seek to drive more value and to move from a cost centre to being a value proposition for an organisation, it is important that all knowledge can be used effectively to generate the largest return on investment. The only way to truly achieve this is to adopt smart laboratory technology.

This is a consistent theme throughout the entire publication. Building a smart laboratory can provide huge benefits to an organisation in terms of increased productivity or value generation, but also through collection management and archiving of data. However, in order to make the most of the investment in 'smart' technologies, it is imperative that a strategy is devised that can look at the needs to the lab and its users to properly adapt and configure the technology accordingly.

Technology will not do the thinking for us, but if properly constructed a smart laboratory can add considerable value. While this guide cannot provide all the answers, it does provide an introduction to everyone that faces the challenge of increasing productivity and data integrity for the modern laboratory workflow.

The authors of the guide are:

Peter Boogaard
Industrial Lab Automation

Siri Segalstad
Segalstad Consulting AS

Joe Liscouski
Institute for Laboratory Automation

Charlie Sodano
eOrganizedWorld

John Trigg
phaseFour Informatics Ltd

Isabel Muñoz-Willery
NL42 Consulting SL

Roberto Castelnovo
NL42 Consulting SL

Cover image and all other images: Shutterstock.com

Building a Smart Laboratory is published by Europa Science, the publishers of *Scientific Computing World* (ISSN 1356-7853). ©2018 Europa Science Ltd. 4 Signet Court, Cambridge, CB5 8LA, UK. All images Shutterstock.com

Design: Zöe Andrews

Tel: +44 (0)1223 221033. Fax: +44 (0)1223 213385.

www.scientific-computing.com/BASL2018

AN INTRODUCTION TO

Building a Smart Laboratory 2018

It's rare for a company to start with a clean slate when making decisions about laboratory automation

This chapter serves as an introduction to this guide *Building a Smart Laboratory 2018*. We hope to highlight the importance of adopting smart laboratory technology but also to guide users through some of the challenges and pitfalls when designing and running the latest technologies in the lab.

For any laboratory a cost/benefit analysis needs to consider the functionality already provided by legacy applications – as well as business justifications. This guide will help you understand what informatics processes are needed in laboratories, and why the laboratory should not merely be seen as a necessary cost centre.

Only by becoming smart – as this guide outlines – can lab managers change that mind-set and generate true value for their organisation.

Many laboratory operations are still predominantly paper-based. Even with the enormous potential to reduce data integrity for compliance, to make global efficiency gains in manufacturing and to increase knowledge sharing, the barriers to implementing successful electronic integrated processes often remain a bridge too far.

The informatics journey

The journey starts with data capture, data processing, and laboratory automation. When samples are being analysed, several types of scientific data are being created. They can be categorised in three different classes.

Raw data refers to all data on which decisions are based. Raw data is created in real-time from an instrument or in real-time from a sensor device.

Metadata is 'data about the data' and it is used for cataloguing, describing, and tagging data resources. It adds basic information, knowledge, and meaning. Metadata helps organise electronic resources, provide digital identification, and helps support archiving and preservation of the resource.

Secondary or processed data describes how raw data is transformed by using scientific methodologies to create results. To maintain data integrity, altering methods to reprocess will require a secured audit trail functionality, data and access security. If metadata is not captured, the ability to find and re-use previous knowledge from scientific experiments is eliminated.

Paperless or less paper?

Data-intensive science is becoming far more mainstream; however, going digital in the laboratory has been a relatively slow process. More than 75 per cent of laboratory analysis starts with a manual process such as weighing; the majority of results of these measurements are still written down or re-typed.

There are exceptions: probably the best example of integrated laboratory automation can be found in how chromatography data handling systems (CDS) operate in modern laboratories. The characteristics of such a system include repeatable, often standardised, automated processes that create a significant amount of raw and processed data.

The paper versus paperless discussion is as old as the existence of commercial computers. In the 1970s, just after the introduction of the first personal computer, Scelbi (Scientific, Electronic and Biological), *Business Week* predicted that computer records would soon completely replace paper. It took at least 35 years before paperless operations were accepted and successfully adopted in many work operations. Although they have been accepted in banking, airlines, healthcare, and retail, they lag behind in science.

The journey from paper to electronic begins

with the transition from paper to digital, which includes both the transfer of paper-based processes to 'glass' and the identification and adoption of information and process standards to harmonise data exchange.

Think exponential

Traditional mainstream LIMS will face challenges. LIMS has been a brilliant tool to manage predictable, repeatable planned sample, test and study data flows, creating structured data generated by laboratories. In R&D environments, unpredictable workflows creating massive amounts of unstructured data showed that current LIMS systems lack the capability effectively to manage this throughput. ELNs are great tools to capture and share complex scientific experiments, while an underlying scientific data management system (SDMS) is used to manage large volumes of data seamlessly.

Data consumer vs data creator examples

For the researcher, the ability to record data, make observations, describe procedures, include images, drawings and diagrams and collaborate with others to find 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 well received.

Product innovation and formulators will need the capability to mine data across projects, analytical methods or formulations to create valuable insights. Transforming unstructured scientific experimental data into a structured equivalent will be mandatory to perform these tasks.

Organisations with a strong consumer marketing focus deal with data mining techniques providing clear pictures of products sold, price, competition and customer demographics.

New trends

The power of life cycle process improvement

The scientist is no longer in the laboratory, but integrated in the overall quality process. Quality should be built into the design throughout the specification, design, and verification process. Performance metrics on non-conformance tracking are mandated and monitored by regulatory authorities. Integrating laboratory systems will add significant value by decreasing non-conformance.

New budgeting and licensing models

Managing operating budgets will be redefined in the next decade. The days of purchasing software as a capital investment (CAPEX) are changing to a new model based upon a 'pay-as-you-go' or philosophy (OPEX). CRM applications such as SalesForce.com started this business model in the traditional enterprise business software segment. Popular applications such as Photoshop, Microsoft Office 365 and Amazon are following these trends rapidly. It is expected that scientific software suppliers will be forced to follow the same model in the years to come. Community collaboration and social networking is changing the value of traditional vendor help desks.

Reduce and simplify workflow complexities

The need to simplify our scientific processes will have a significant impact on reducing data integrity challenges. For example, balance and titrator instruments may store approved and pre-validated methods and industry best practice workflows in their firmware.

Adopt and use industry standards and processes

Initiatives such as the Allotrope Foundation are working hard to apply 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.

Consolidation and harmonisation of systems

Most laboratories already depend on an informatics hub comprising one or more of the major tools: laboratory information management systems (LIMS); electronic laboratory notebooks (ELN); scientific data management systems (SDMS); chromatography data-handling systems (CDS) and laboratory execution systems (LES). The trend over recent years has been towards convergence, applying best practice industry standard processes to harmonise multisite deployments. Cost reduction to interface harmonised processes to ERP (SAP), MES and CAPA results in lower maintenance and validation costs with a significant overall higher system availability for end-users.

Mobile computing

While many other industries are implementing modern tools to connect equipment wirelessly, many laboratories still write scientific results on a piece of paper, or re-type them into a computer or tablet. Many modern ELN and LES systems allow electronic connection to a (wireless) network. However, to integrate simple instruments like a pH balance, titration and Karl-Fischer instruments to mobile devices, a simpler approach is required in order to achieve mainstream adoption. The acceptance of tablets and mobile devices will expand exponentially in the laboratory.

Laboratories will need to manage the challenges presented by new consumers of scientific data outside traditional laboratory operations. Non-invasive, end-to-end strategies will connect science to operational excellence. Technology will be critical, but our ability to change our mind-set to enable this cross-functional collaboration will be the real challenge. ■

Adapting to change

Much of the change that drives new processes or methods in the laboratory is based on regulation from that aims to more tightly control the way in which data is collected, stored and handled.

Many laboratory users will be aware of previous regulations such as Title 21 CFR Part 11, part of Title 21 of the Code of Federal Regulations that establishes the United States Food and Drug Administration (FDA) regulations on electronic records and electronic signatures (ERES).[1]

Part 11, of the document, as it is commonly called, defines the criteria under which electronic records and electronic signatures are considered trustworthy and equivalent to paper records.

However new regulation around General Data Protection Regulation (GDPR) and data integrity are new standards that laboratory users must now familiarise themselves with. For many users GDPR will not be applicable as it only relates patient data or companies that hold data of EU citizens. However, if in a clinical setting GDPR could have a huge effect on the way that you store patient data.[2]

In addition to GDPR lab managers must also familiarise themselves with pending regulation on Data Integrity (DI) which hopes to improve completeness, consistency, and accuracy of data recorded by laboratories[3]. In simple terms this means abiding by principles such as ALCOA (attributable, legible, contemporaneous, original, and accurate). However it is advised that lab managers and users explore the ramifications of this new regulation to see how it might affect daily workflows.

References

1. https://www.fda.gov/regulatoryinformation guidances/ucm125067.htm
2. https://www.ncbi.nlm.nih.gov/pmc/articles PMC5346164
3. https://www.fda.gov/downloads/drugs/guidances ucm495891.pdf

'eConnect, eDecide, eManage, eArchive'

The key layers of a laboratory paperless strategy

Isabel Muñoz-Willery and Roberto Castelnovo, of NL42 Consulting, highlight the importance of developing a robust strategy for the adoption paperless laboratory operations

In the new era of the internet of things and artificial intelligence, the majority of laboratories still have a long way to move from paper-based processes to paperless ones.

The electronic data life cycle, as it is described in several regulations and documents used in paperless projects, can be divided in four layers of data, information and activities: eConnect; eManage; eDecide and eArchive.

These keywords refer to initial capture of data, the data management to create useful information, the decisions taken based on information and data available in the lower layers and, finally, the electronic data archiving to ensure long-term availability of the information and the related data.

Those are the four-main streams that will be discussed in detail at the Paperless Lab Academy 2018. The annual European event aims to become a learning platform for anyone looking to consolidate, integrate or simplify their data management systems.

'eConnect': effective workflows based on self-documenting data capture strategies

Even if data integrity is a critical aspect of the entire data life cycle, data capture requires a strong focus from both the inspectors and auditors. Most lab instruments are now offered with intelligent software embedded into them. Labware and sensors are beginning to embrace the internet of things, ensuring the collection of the raw data and the related metadata which can then be transferred to the next phase of the data life cycle.

Several laboratories are using instruments which are not able to connect the current platforms. While searching for the business justification for their replacement, intermediate solutions should be considered to generate digital inputs and reduce paper-based processes and manual transcriptions. The goal is to reduce the manual documentation, the risk of human errors, and more importantly, to maintain the information about the source that has generated the raw data.

The raw data may be a critical part of the activities performed in the systems of the upper layers. Data management and the creation of meaningful information and decisions should be always taken with the possibility to go back to the original data from the system in which it was generated.

Finally, while in this first stage of collecting data we should not obviate the ones coming from collaborators. Collaborators are generators of data and potential sources of information. If external organisations such as academic contributors or outsourced services from CRO and CMO are generating the data, it can create immediate security concerns. With the latest GDPR considerations, we need to incorporate data protection assessment at least on the most vulnerable data. By May 2018, companies will need to design their processes and also include serious considerations on cybersecurity protection to avoid any risk in losing data.

LABWARE 7

LIMS and ELN together in a single integrated software platform.
A laboratory automation solution for the entire enterprise.

LABWARE
Results Count

Offices worldwide supporting customers
in more than 100 countries

www.labware.com

'eManage': generation of meaningful information from trusted data

The 'translation' from data to information is the key principle of this layer of activities typically performed in the most well-known systems The real challenge in the new era of Internet of Lab Things (IoLT) is not about picking up the right acronym for the lab. The challenge lies in identifying the right solutions that provide answers to a series of requirements: secured connectivity without large investment; usability with limited customisation; ability to share information using the newest technologies; mobile devices and web-access without performing complex platform implementations; and the possibility to use the software as a service.

We are observing a large market transformation in this area. The presence of systems which are offering a large set of functionalities and product offerings based on new technologies.

Multiple software modules adapted to specific laboratory activities and software platforms allow the creation of personalised solutions with no need to customise but rather configure the system to the needs of the user.

This revolution will generate large benefits for the laboratories because the selection of the solutions will be based on the needs rather than the capabilities.

These modules should respond to a few critical requirements in order to become part of the 'solution': easily connectable to the 'eConnect' layer; easily connectable to modules of the 'eManage' layer; easily accessible from browsers and mobile devices; and easily accessible from the 'eDecide' layer.

What is the end goal?

On one hand, the final goal should be to interface the 'solution' with the multiple generators of raw data in order to enable the review directly at the source at any time. Additionally, the possibility to exchange information between the modules of the 'eManage' layer, in a flawless manner, should allow the access and interpretation of all data to generate meaningful information.

The possibility to access the 'modules' from any remote location or even from mobile devices in order to manage all the information in the shortest period of time. The possibility to provide aggregated information to the next layer of systems where decisions are taken.

Is this real? Absolutely. The technology has evolved to the level that all these goals could be reached.

Numerous solutions are already implemented in various markets where they are using the newest technologies. The laboratory informatics systems will have to be ready for this new era too.

'eDecide': Rapid decisions taken from meaningful information

In the everyday activities of a laboratory, we are getting used to perform them very rapidly and decisions should be taken in short time. Little remains available for data review, approval of data and creation of related documents. The request coming from laboratory's customers, both internal and external is a prompt answer.

The removal of manual processes, of paper-based activities and mix of information sitting in different systems is essential for taking faster decisions. Only paperless processes shorten the periods of review of the information and ease rapid decision-making which can then be communicated immediately to the relevant stakeholders. New approaches like the review by exception are helping to increase the efficiency of this process.

The laboratories that are able to respond to these requests on time and with the adequate level of quality will transform from cost centers to value generators.

Decisions should be taken according to the available information. Today many software providers offer simple tools presenting the information in a graphical view, showing the outliers, highlighting the areas of attention, allowing the 'drill-down' approach when needed.

Fact is that solutions providers, integrators and customers are joining efforts in organisations like the allotrope foundation, Pistoia Alliance, SiLA consortium to consolidate outputs and tools, that could one day lead to the creation of one single user interface, one single way of showing the information in a unique and personalised dashboard.

Simple reports created automatically overnight and available in the 'eDecide layer' first thing in the morning. A new 'control room' of the laboratory where decisions are taken to correct situations not in line with the expectations, where scheduling changes are adjusted to ensure that the activities are completed on time, on budget and according to the customer expectations.

Is this real? Yes, again. Great reporting and business intelligence tools are now available to integrate the information coming from different systems and present in a simple and graphical way. All what the managers need at their fingertips. Moreover, these tools are able to dig into the underlying systems to view the information and related raw data, when needed.

We will finally see one single screen open on the computers of the managers instead of multiple windows jumping from one system to another in order to desperately collect all the necessary information required in a given moment for a given decision to be taken urgently.

'eArchive': essentials to secure long-term multi-departmental archiving

A key objective in operating with efficient archival approach is to reduce the challenge of finding the right data. Considering the growing digital universe, archiving can no longer be left behind in a project and considered only once it is too late. Nowadays, we often hear about concerns on legibility and format consistency along the time for a given retention time that might end up requiring access to obsolete technologies.

Archiving should be approached and designed to reduce multiple types of risk: knowledge limited to one critical person, security and loss of data.

A comprehensive archiving protocol should eliminate the struggle to find the data to the point of desperately looking for the person owning the knowledge of where it is.

A corporate master data management and vocabulary model should support a correct management and archival, facilitating a flawless track record of the data.

During the Paperless Lab Academy 2018, several presentations will focus on this item that too often is approached too late in a 'paperless' project. The archiving strategy requires a clear definition of the business requirements and, also the potential technical challenges.

The ability to archive and then retrieve unstructured data is becoming an urgent need which must be solved in R&D laboratories. Solutions providers are dedicating resources to this matter and positioning their data management software to address the need for better archiving and retrieval. Above all, the 'eArchive' strategy is one that requires stronger alignment within the whole company in order to build up a reference master data management strategy at an enterprise-level. ■

The removal of manual processes, of paper-based activities and mix of information sitting in different systems is essential for taking faster decisions

"

dotmatics
knowledge solutions

Scalable Scientific Informatics Solutions

Dotmatics' comprehensive scientific solutions include tools for knowledge management, data storage, enterprise querying and reporting, and data analysis and visualization for biopharma and scientific organizations.

dotmatics.com

Dealing with data

Informatics experts share their experiences on the implementing new technologies and manging change in the modern laboratory

Mark Gonzalez
Technical director at Labware

What technologies are requested by laboratory users?

Mark highlighted that there are clear divisions between the two primary groups of existing customers and potential users.

'In terms of technology the question that existing users are asking about most often is mobile. That is not to say that they have a clear plan on how to use the technology but they have smart phones and tablets in their personal lives and they want to know what we could do using that layer,' said Gonzalez. 'That is the technical question that we get the most from our existing users. They don't tend to ask about cloud because they have a running system. The IT department might be interested in moving to the cloud but since they have the system already running and they are not likely to want to change that in the short term,' Gonzalez added.

Gonzalez noted that mobile technology as a solution for laboratory users 'is a solution that needs to solve real-life problems.'

'What we want to do is solve the right problems we don't want to just throw out a bunch of technology that doesn't really solve anything of any business value.'

One example that he noted was the ability to use mobile devices in untethered mode. This could allow users to perform actions such as entering data without a continuous connection to the LIMS server. Once the connection is re-established the data can be automatically sent to the LIMS system. 'One value of mobile technology is that people could work remotely to collect data, even if they don't have a connection to the LIMS server,' concluded Gonzalez

Renaud Acker
Chief operating officer at Agilab

What are the main challenges that your users face when deploying digital informatics technology?

Renaud Acker explains that, for many of AgiLab's customers, 'change control' is the main challenge: 'Processes have changed by using a new generation of software. Users must be trained, standard operating procedures (SOPs) must be adapted, data handling and traceability must be managed in a different way.

'This means that lab software should be user-friendly for daily use. Screens must be clear with adapted vocabulary,' stated Acker. 'However, it must also be adapted to laboratory processes and objectives in order to increase efficiency and productivity and finally it must facilitate collaboration between scientists.

'The main issue is to handle data, not only to store it but also to be able to use that data effectively,' stressed Acker. 'Laboratories are producing and accumulating more and more data from experiments, analysis, bibliography and other areas. For instance, one screening campaign could generate hundreds of thousands of results, a query on a citation source like Pubmed can report thousands of references.

'The challenge is to centralise data, to manage and gather information, to generate knowledge from data – and to keep track of what has been done, how it has been done, if it has worked or not. Big data technologies will be very useful to annotate, explore and exploit the whole set of data generated in labs and gathered from external public sources.

While there are clear benefits to using the latest software, cost of investment can be a big issue that prevents companies from replacing legacy infrastructure – but it is not the only reason, as Acker explains.

'There are at least two main reasons why labs don't move easily to new lab software. Many companies and labs have spent fortunes in their first generation of lab software. Secondly, they have customised these products with considerable effort and money, so they aren't eager to move. When a Lab needs to be compliant to GMP, GLP, etc it has many other points to manage: change control, system validation, certification and audits.'

All of these aspects can make a move more challenging, but ultimately choosing not to upgrade impacts agility – and the speed and quality of further laboratory operations.

Another aspect that AgiLab was keen to stress was that cloud deployments are increasingly seen as a good choice for many laboratories. However, the move to cloud based informatics requires a user to change their mindset as they move from silos of data to a more fluid model of shared data sets and collaboration.

'Labs are still working in silos,' added Acker. 'New R&D processes should break this logic in order not only to exchange data but mostly to anticipate issues by gathering scientists working on a project. Collaboration is essential for R&D project success. Cloud applications could help to exchange data and ideas between labs in different locations, between industrial, partners and academics.'

Acker concluded that cloud-based laboratory informatics is growing due to a number of factors including their robust security, the potential for hosting management of services off-premise and the use of cloud subscription models that can reduce initial investment and running costs.

Oscar Kox
Business delopment manager at Ivention

How important are digital technologies to the modern laboratory?

'There is a lot of innovation available in the market but I don't think many labs are picking it up as early adopters,' said Kox.

'People should ask themselves how important is it adopt new technologies – to innovate in the lab. Having worked in this industry for more than 20 years – of course it is important. You want to see new technology getting into the laboratory either because you want to reduce FTE, you want to increase throughput or improve quality.'

Kox gave an example of large implementation that iVention is managing in Europe that is consolidating as many as seven individual implementations with their own custom software, with additional software connected to it.

'They cannot upgrade everything all at once,' he said.

The presence of custom software in each implementation means that each installation is essentially a new piece of software.

'Now if you compare this to the capabilities of a web-based system you can rollout to all of those sites without custom software – there is a big benefit,' said Kox.

'If there is a LIMS project that people who are now looking for a new LIMS or ELN, the decision they make now will affect them for the coming five to 10 years, because that is the investment that you are looking at.'

Kox stressed customers should ask themselves: will this big conventional LIMS vendor help me to innovate? 'That is where the gap comes in. There's a lot of innovation out there but can I adopt it right now, because of the systems I have in place?'

He explained that iVention has installed systems across very large organisations. He gave an example of a pharma client who wanted to roll out a system for 300 users across seven countries, over eight months. Cox also mentioned that this solution was hosted for the client by iVention.

'I don't think there are many of those rollouts completed successfully with a conventional LIMS system,' said Kox.

'They are a big company with their own IT department and we are hosting it for them because we have all the technology in place to automate everything, so all the upgrades can be done automatically.' He explained the success of this rollout has meant this company is now using iVention as a strategic partner for much larger rollouts in the future.

Kox said: 'I have seen organisations with very old software, which can be costly and time consuming to maintain and upgrade. Some IT directors would say the upgrade would cost more than the original installation, so they either try and run for a few more years or select a new system.'

He said one of the main challenges when dealing with legacy LIMS or ELN systems is a lack of maintenance and upgradability: 'The biggest thing I see is customers paying maintenance and they cannot upgrade. Support cannot help them because they have an old version and in many cases this support money is wasted because the system is too old to be properly supported.

'I would strongly recommend firms look at their maintenance contacts and ask themselves "what are we getting back from it?"' ■

The smart laboratory

This chapter discusses what we mean by a 'smart laboratory' and its role in an integrated business. We also look at the development of computerised laboratory data and information management; the relationships between laboratory instruments and automation (data acquisition); laboratory informatics systems (information management); and higher-level enterprise systems and how they align with knowledge management initiatives.

The progressive 'digitisation' of the laboratory offers an unprecedented opportunity not only to increase laboratory efficiency and productivity, but also to move towards 'predictive science', where accumulated explicit knowledge and computer algorithms can be exploited to bring about greater understanding of materials, products, and processes

Today the landscape for laboratory technologies is broad and varied. This is true purely in terms of the variation of management systems and other software packages but also due to the proliferation of additional technology such as cloud, mobile technologies and more recently the IoT.

There is no specific definition of a 'smart laboratory'. The term is often used in different contexts to imply either that a laboratory is designed to optimise its physical layout, that it incorporates the latest technology to control the laboratory environment, or that the laboratory is using the latest technology to manage its scientific activities. For the purposes of this publication, it is the latter definition that applies.

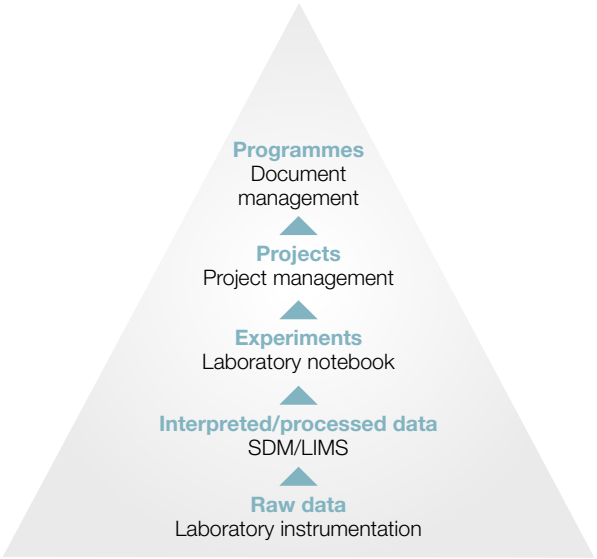
Using technology to manage scientific endeavours is conceptually a straightforward task but the subtlety lies in choosing the right combination of technologies that can be adapted to suit the use case of a specific laboratory which may be dictated by geography and personnel as much as it is driven by the availability of technology. As such the right answer to setting up a smart laboratory is not to adopt all possible technological features but to identify which areas of the laboratory need to be accelerated or improved upon.

A simple example of this could be found in a common problem facing many laboratories – data generated through 'dumb' instrumentation such as pH meter or weighing scales. Instruments that are not connected directly to a (Laboratory Informatics Management System) LIMS or Electronic Laboratory Notebook (ELN) type management system present opportunities to introduce error through human data entry but there are multiple ways to solve this problem.

One would be to buy new scales for example. Purchasing a new instrument with smart capabilities could feed that data directly into the LIMS reducing the chance for error. Another approach would be the use of mobile devices which could be used to capture the data at the bench another would be to use a raspberry Pi like device connected to the internet to take the result and feed it into the LIMS. The choice around whether mobile, IoT or new instruments is one that can only be answered on a case by case basis – there is no one size fits all solution for every laboratory.

The introduction of industrial R&D laboratories heralded a new era of innovation and development dependent on the skills, knowledge and creativity of individual scientists. The evolution has continued into the 'information age' with a growing dependence on information technology, both as an integral part of the

FIG 1 Information structure



scientific process, and as a means of managing scientific information and knowledge.

Laboratory information has traditionally been managed on paper, typically in the form of the paper laboratory notebook, worksheets and reports. This provided a simple and portable means of recording ideas, hypotheses, descriptions of laboratory apparatus and laboratory procedures, results, observations, and conclusions. As such, the lab notebook served as both a scientific and business record. However, the introduction of digital technologies to the laboratory has brought about significant change.

From the basic application of computational power to undertake scientific calculations at unprecedented speeds, to the current situation of extensive and sophisticated laboratory automation, black box measurement devices, and multiuser information management systems, technology is causing glassware and paper notebooks to become increasingly rare in the laboratory landscape. The evolution of sophisticated lab instrumentation, data and information management systems, and electronic record keeping has brought about a revolution in the process of acquiring and managing laboratory data and information. However, the underlying principles of the scientific method are unchanged, supporting the formulation, testing, and modification of hypotheses by means of systematic observation, measurement, and experimentation. In our context, a smart laboratory seeks to deploy modern tools and technologies to improve the efficiency of the scientific method by providing seamless integration of systems, searchable repositories of data of proven integrity, authenticity and reliability, and the elimination of mindless and unproductive paper-based processes.

At the heart of the smart laboratory is a simple model (see Figure 1) that defines the conceptual, multi-layered relationship between data, information, and knowledge.

The triangle represents the different layers of abstraction that exist in laboratory workflows. These are almost always handled by different systems. The 'experiment' level is the focal point for cross-disciplinary collaboration: the point at which the scientific work is collated and traditionally handled by the paper laboratory notebook.

Above the experimental layer is a management context that is handled by established groupware and document management tools at the 'programme' level, and by standard 'office' tools at the project level. Below the experiment level there is an increasing specialisation of data types and tools, typically encompassing laboratory instrumentation and multi-user sample and test management systems. The triangle also represents the transformation of data to knowledge, the journey from data capture to usable and reusable knowledge that is at the heart of the smart laboratory.

The introduction of ELNs therefore opens up the possibility of a more strategic approach, which, in theory at least, offers the opportunity for an integrated and 'smart' solution.

A frequently articulated fear about the relentless incorporation of technology in scientific processes is the extent to which it can de-humanise laboratory activities and reduce the demand for intellectual input, or indeed, any fundamental knowledge about the science and technology processes that are in use. The objective of this publication is to present a basic guide to the most common components of a 'smart laboratory', to give some general background to the benefits they deliver, and to provide some guidance to how to go about building a smart laboratory.

The two primary areas of technology that apply to a smart laboratory can be broadly categorised as laboratory automation and laboratory informatics. In general, laboratory automation refers to the use of technology to streamline or substitute manual manipulation of equipment and processes. The field of laboratory automation comprises many different automated laboratory instruments, devices, software algorithms, and methodologies used to enable, expedite, and increase the efficiency and effectiveness of scientific research in labs. Laboratory informatics generally refers to the application of information technology to the handling of laboratory data and information, and optimising laboratory operations.

In practice, it is difficult to define a boundary between the two 'technologies' but, in the context of this publication, chapter three (Data) will provide an overview of laboratory instrumentation and automation, predominantly data capture.

Chapter four (Information) will look at the four major multi-user tools that fall into the 'informatics' category, identifying their similarities, differences and the relationship between them. Chapters three and four, therefore, focus on the acquisition and management of data and information, whereas chapter five (Knowledge) will provide guidance about the long-term retention and accessibility of laboratory knowledge through online storage and search algorithms that aim to offer additional benefits through the re-use of existing information, the avoidance of repeating work, and enhancing the ability to communicate and collaborate.

The underlying purpose of laboratory automation and laboratory informatics is to increase productivity, improve data quality, to reduce laboratory process cycle times, and to facilitate laboratory data acquisition and data processing techniques that otherwise would be impossible. Laboratory work is, however, just one step in a broader business process – and therefore, in order to realise full benefit from being 'smart', it is essential that the laboratory workflow is consistent with business requirements and is integrated into the business infrastructure in order for the business to achieve timely progress and remain competitive.

Chapter seven (Beyond the laboratory) will examine the relationship between laboratory processes and workflows with key business issues such as regulatory compliance and patent evidence creation, and will also address productivity and business efficiency.

Chapter eight (Practical considerations in specifying and building the smart laboratory) is therefore devoted to the process of making the laboratory 'smart', taking into account the functional needs and technology considerations to meet the requirements of the business, and addressing the impact of change on laboratory workers. ■

DATA

Instrumentation

This chapter will consider the different classes of instruments and computerised instrument systems to be found in laboratories and the role they play in computerised experiments and sample processing – and the steady progress towards all-electronic laboratories.

However, the choice of best-of-breed laboratory instruments and instrument systems can present challenges when it comes to getting everything to work together in a seamless way. The final part of this chapter will look at the issue of standard data interchange formats, the extent of the challenge, and some of the initiatives to address them

Simple laboratory instruments

Devices such as analytical balances and pH meters use low-level processing to carry out basic functions that make them easier to work with. The tare function on a balance avoids a subtraction step and makes it much easier to weigh out a specific quantity of material.

Connecting them to an electronic lab notebook (ELN), a laboratory information management system (LIMS), a lab execution system (LES), or a robot, adds computer-controlled sensing capability that can significantly off-load manual work. Accessing that balance through an ELN or LES permits direct insertion of the measurement into the database and avoids the risk of transcription errors. In addition, the informatics software can catch errors and carry out calculations that might be needed in later steps of the procedure.

The connection between the instrument and computer system may be as simple as an RS-232 connection or USB. Direct Ethernet connections or connections through serial-to-Ethernet converters can offer more flexibility by permitting access to the device from different software systems and users. The inclusion of smart technologies in instrumentation significantly improves both their utility and the labs' workflow.

Computerised instrument systems

The improvement in workflow becomes more evident as the level of sophistication of the software increases. It is rare to find commercial instrumentation that doesn't have processing capability either within the instruments' packaging or, through a connection to an external computer system.

The choice of dedicated computer-instrument combinations vs. multi-user, multi-instrument packages is worth careful consideration. The most common example is chromatography, which has options from both instrument vendors and third-party suppliers.

One of the major differences is data access and management. In a dedicated format, each computer's data system is independent and has to be managed individually, including backups to servers.

It also means that searching for data may be more difficult. With multi-user/instrument systems there is only one database that needs to be searched and managed.

If you are considering connecting the systems

to a LIMS or ELN, make the connections as simple as possible. If an instrument supported by the software needs to be replaced, changing the connection will be simpler.

Licence costs are also a factor. Dedicated formats require a licence for each system. Shared-access systems have more flexible licensing considerations. Some have a cost per user and connected instrument; others have a cost per active user/instrument schedules.

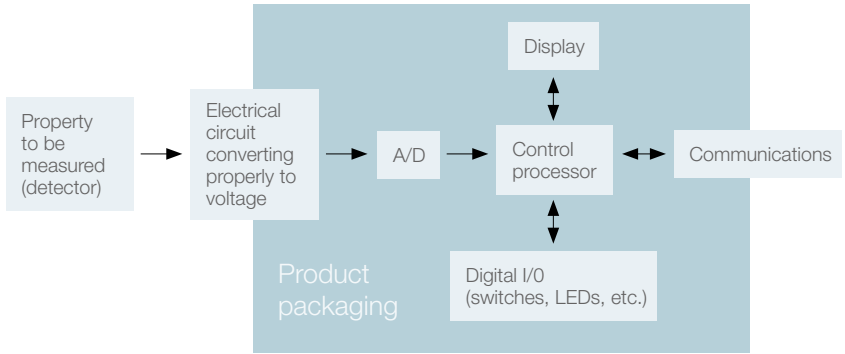
In the latter case, there are eight instruments and four analysts, of which only half may be simultaneously active, licenses for only four instruments and two users are needed.

One factor that needs attention is the education of laboratory staff in the use of computer-instrument systems.

While instrument software systems are capable of doing a great deal, their ability to function is often governed by user-defined parameters that affect, at least in chromatography, baseline-corrections, area allocation for unresolved peaks, etc. Carefully adjusted and tuned parameters will yield good results, but problems can occur if they are not managed and checked for each run.

Type of A/D	Capability
Successive approximation	These are general-purpose devices suitable for a wide range of applications. They have limited resolution, but have amplifiers for low-level signals, and can sequentially access multiple input channels. Their resolutions are up to 18 bits (262,144 steps) and sampling speeds of up to five million samples per second (sps). The higher the resolution, the slower the sampling speed.
Integrating	Good for low speed sampling (<100 sps), high resolution >14 bits, single channel inputs, with good noise rejection. Often used in chromatography.
Sigma-Delta A/D	Up to 24 bits of resolution, single channel input – may not be efficient for multi-channel inputs, low speed, may replace integrating A/Ds.
Flash	Single channel input, 8-bit conversion, approximately 1 billion SPS. Good for very high-speed applications, where low resolution is not a problem. You can digitise electrical noise.

FIG 2 Analogue data acquisition



Instrument data management

The issue of instrument data management is a significant one and requires considerable planning. Connecting instruments to a LIMS or ELN is a common practice, though often not an easy one if the informatics vendor hasn't provided a mechanism for interfacing equipment. Depending on how things are set up, only a portion of the information in the instrument data system is transferred to the informatics system.

If the transfer is the result of a worklist execution of a quantitative analysis, only the final result may be transferred – the reference data still resides on the instrument system. The result is a distributed data structure. In regulated environments, this means that links to the backup information have to be maintained within the LIMS or ELN, so that it can be traced back to the original analysis.

The situation becomes more interesting when instrument data systems change or are retired. Access still has to be maintained to the data those systems hold. One approach is virtualising the instrument data system so that the operating system, instrument support software, and the data are archived together on a server. (Virtualisation is, in part, a process of making a copy of everything on a computer so that it can be stored on a server as a file or 'virtual container' and then executed on the server without the need for the original hardware. It can be backed up or archived, (so that it is protected from loss). In the smart laboratory, system management is a significant function – one that may be new to many facilities. The benefits of doing it smartly are significant.

Computer-controlled experiments and sample processing

Adding intelligence to lab operations isn't limited to processing instrument data, it extends to an earlier phase of the analysis: sample preparation. Robotic systems can take samples – as they are created – and transfer the format to that needed by the instrument. Robotic arms – still appropriate for many applications – have been replaced with components more suitable to the task, particularly where liquid handling is the dominant activity, as in life science applications.

Success in automating sample preparation depends heavily on thoroughly analysing the process in question and determining:

- Whether or not the process is well documented and understood (no undocumented short-cuts or workarounds that are critical to success), and whether improvements or changes can be made without adversely impacting the underlying science;
- Suitability for automation: whether or not there are any significant barriers (equipment, etc.) to automation and whether they can be resolved;
- That the return in investment is acceptable and that automation is superior to other alternatives such as outsourcing, particularly for shorter-term applications; and
- That the people implementing the project have the technical and project management skills appropriate for the work.

The tools available for successfully implementing a process are clearly superior to what was available in the past. Rather than having a robot adapt to equipment that was made for people to work with, equipment has been designed for automation – a major advance. In the life sciences, the adoption of the microplate as a standard format multi-sample holder (typically 96 wells, but can have 384 or 1,536 wells – denser forms have been manufactured) has fostered the commercial availability of readers, shakers, washers, handlers, stackers, and liquid additions systems, which makes the design of preparation and analysis systems easier. Rather than processing samples one at a time, as was done in early technologies, parallel processing of multiple samples is performed to increase productivity.

Another area of development is the ability to centralise sample preparation and then distribute

the samples to instrumentation outside the sample prep area through pneumatic tubes. This technology offers increased efficiency by putting the preparation phase in one place so that solvents and preparation equipment can be easily managed, with analysis taking place elsewhere. This is particularly useful if safety is an issue.

Across the landscape of laboratory types and industries, the application of sample preparation robotics is patchy at best. Success and commercial interest have favoured areas where standardisation in sample formats has taken place.

The development of microplate sample formats, including variations such as tape systems that maintain the same sample cell organisation in life sciences, and standard sample vials for auto-samplers, are common examples. Standard sample geometries give vendors a basis for successful product development if those products can have wider use rather than being limited to niche markets.

Building a smart laboratory needs to look beyond commonplace approaches and make better use of the potential that exists in informatics technologies

“

Putting the pieces together

It's not enough to consider in isolation sample preparation, the introduction of samples into instruments, the instruments themselves, and the data systems that support them. Linking them together provides a train of tasks that can lead to an automated sample processin system as shown in Figure 3.

The control/response link is needed to synchronise sample introduction and data acquisition. Depending on the nature of the work, that link can extend to sample preparation. The end result is a system that not only provides higher productivity than manual methods, but does so with reduced operating costs (after the initial development investment).

However, building a smart laboratory needs to look beyond commonplace approaches and make better use of the potential that exists in informatics technologies. Extending that train of elements to include a LIMS, for example, has additional benefits. The initial diagram above would result in a worklist of samples with the test results that would be sent to a LIMS for incorporation into its database.

Suppose there was a working link between a LIMS and the data system that would send sample results individually, and that each sample processed by the instrument would wait until the data system told it to go ahead. The LIMS has the expected range for valid results and the acceptable limits. If a result exceeded the range, several things could happen:

- The analyst would be notified;
- The analysis system would be notified that the test should be repeated;
- If confirmed, standards would be run to confirm that the system was operating properly; and
- If the system were not operating according to SOPs, the system would stop to avoid wasting material and notify the analyst.

The introduction of a feedback facility would significantly improve productivity.

At the end of the analysis, any results that are outside expected limits would have been checked and the systems integrity verified. Making this happen depends on connectivity and the ability to integrate components.

Instrument integration

In order for the example described above to work, components must be connected in a way that permits change without rebuilding the entire processing train from scratch. Information technology has learned those lessons repeatedly as computing moved from proprietary products and components to user friendly consumer systems.

Consumer level systems aren't any less capable than the earlier private-brand-only systems, they are just easier to manage and smarter in design.

Small Computer Systems Interconnect, Firewire and Universal Serial Bus are a few examples of integration methods that enabled the user to extend the basic capability and have ready access to a third-party market of useful components. It also allowed the computer vendors to concentrate on their core product and satisfy end-user needs through partnerships; each vendor could concentrate on what they did best and the resulting synergy gave the users what they needed.

Now these traditional methods are being surpassed by the IOT or wireless connected devices but the argument for connecting devices still remains the same – is the value added worth the investment? The answer depends on the instrument, but generally it is more effective to connect the most widely used instruments such as PH meters and weighing scales.

Connections are only part of the issue. The more significant factor is the structure of the data that is being exchanged: how it is formatted; and the organisation of the content. In the examples above, that is managed by the use of standard device drivers or, when called for, specialised device handlers that are loaded once by the user.

In short, hardware and software are designed for integration, otherwise vendors find themselves at a disadvantage in the marketplace.

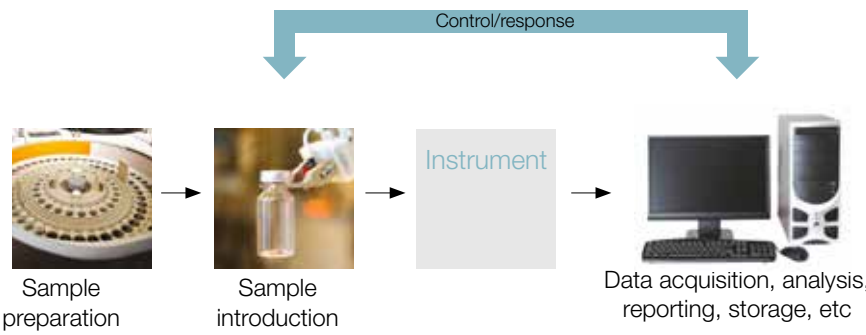
Laboratory software comes with a different mindset. Instrument support software was designed first and foremost to support the vendor's instrument and provide facilities that weren't part of the device, such as data analysis. Integration with other systems wasn't a factor.

That is changing. The increasing demand for higher productivity and better return on investment has resulted in the need for systems integration to get overall better systems performance; part of that measure is to reduce the need for human interaction with the system. Integration should result in:

Ease-of-use: integrated systems are expected to take less effort to get things done;
Improved productivity, streamlined operations: the number of steps needed to accomplish a task should be reduced;

- Avoiding duplicate data: no need to look in multiple places;

FIG 3 An automated sample processing system



- Avoiding transcription errors: integration will result in electronic transfers that should\ be accurate; this avoids the need to enter and verify data transfers manually;
- Improving workflow and the movement of lab data: reducing the need for people to make connections between systems – integration facilitates workflow; and
- More cost-effective, efficient lab operations.

The problem of integration, streamlining operations, and better productivity has been addressed via automation before: in manufacturing applications and clinical labs. In the 1980s, clinical lab managers recognised the only way they were going to meet their financial objectives was to use automation to its fullest capability and drive integration within their systems.

The programme came under the title 'Total laboratory automation' and resulted in a series of standards that allowed instrument data systems to connect with Laboratory Information Systems (equivalent to LIMS) and hospital administrative systems.

Those standards were aggregated under HL7 (www.hl7.org), which provides both message and data formatting. While hospital and clinical systems have the advantages of a more limited range of testing and sample types, making standardisation easier, there is nothing in their structure to prevent them being applied to a wider range of instruments, such as mass spectrometry.

An examination of the HL7 structure suggests that it would be a good foundation for solving integration problems in most laboratories.

From the standpoint of data transfer and communications, the needs of clinical labs match those in other areas. The major changes would be in elements, such as the data dictionaries, and field descriptions, which are specific to hospital and patient requirements.

A cross-industry solution would benefit vendors as it would simplify their engineering and support, provide a product with wider market appeal, and encourage them to implement it as a solution.

Most of the early standards work carried out outside the clinical industry has focused on data encapsulation, while more recent efforts have included communications protocols:

- In the 1990s, efforts by instrument vendors led to the development of the andi standards (Analytical Data Interchange) which resulted in ASTM E1947 – 98(2009) Standard Specification for Analytical Data Interchange Protocol for Chromatographic Data, which uses the public domain netCDF data base structure, providing platform independence. This standard is supported in several vendor products but doesn't see widespread use;
- SiLA Rapid Integration (www sila-standard.org). The website states: 'The SiLA consortium for Standardisation in Lab Automation develops and introduces new interface and data management standards, allowing rapid integration of lab automation ystems. SiLA is a not-for-profit membership corporation with a global footprint and is open to institutions, corporations and individuals active in the life science lab automation industry. Leading system manufacturers, software suppliers, system integrators and pharma/biotech corporations have joined the SiLA consortium and contribute in different technical work groups with their highly skilled experts';
- The Pistoia Alliance (www.pistoiaalliance org) states: 'The Pistoia Alliance is a global, not-for-profit precompetitive alliance of life science companies, vendors, publishers and academic groups that aims to lower barriers to innovation by improving the interoperability of R&D business processes. We differ from standards groups because we bring together the key constituents to identify the root causes that lead to R&D inefficiencies and develop best practices and technology pilots to overcome common obstacles';
- The Allotrope Foundation (www allotrope.org): 'The Allotrope Foundation is an international association of biotech and pharmaceutical companies building a common laboratory information framework ('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'; and
- The AnIML markup language for analytical data (animl.sourceforge.net) is developing a standard specification under the ASTM (www astm.org/DATABASE.CART/WORKITEMS WK23265.htm) that is designed to be widely applicable to instrument data. Initial efforts are planned to result in implementations for chromatography and spectroscopy.

A concern with the second, third and fourth points above is that they are primarily aimed at the pharmaceutical and biotech industries. While vendors will want to court that market, the narrow focus may slow adoption since it could lead to the development of standards for different industries, increasing the implementation and support costs. Common issues across industries and applications lead to common solutions.

The issue of integrating instruments with informatics software is not lost on the vendors. Their product suites offer connection capabilities for a number of instrument types to ease the work.

Planning a laboratory's information handling requirements should start from the most critical point, a LIMS for example, and then on to support additional technologies.

In the 1980s, clinical lab managers recognised that the only way they were going to meet their financial objectives was to use automation to its fullest capability and drive integration within their systems

"

Chapter summary

The transition from processing samples and experiment manually to the use of electronic systems to record data is a critical boundary. It moves from working with real things to their digital representation in binary formats. Everything else in the smart laboratory depends on the integrity and reliability of that transformation.

The planners of laboratory systems may never have to program a data acquisition system, but they do have to understand how such systems function, and what the educated lab professional's role is in their use. Such preparatory work will enable the planners to take full advantage of commercial products.

One key to improving laboratory productivity is to develop an automated process for sample preparation, introducing the sample into the instrument, making measurements, and then forwarding that data into systems for storage, management, and use. Understanding the elements and options for these systems is the basis for engineering systems that meet the needs of current and future laboratory work.

The development of the smart laboratory is at a tipping point. As users become more aware of what is possible, their satisfaction with the status quo will diminish as they recognise the potential of better-designed and integrated systems.

Realising that potential depends on the same elements that have been successful in manufacturing, computer graphics, electronics, and other fields: an underlying architecture for integration, based on communications and data encapsulation/interchange standards. ■

INFORMATION

Laboratory informatics tools

This chapter will look at the four major laboratory informatics tools – laboratory information management systems (LIMS), electronic laboratory notebooks (ELNs), laboratory execution systems (LES) and scientific data management systems (SDMS) – their differences and how they relate to each other. Each of these systems functions at or around the 'Information' layer (see Figure 1) and typically serves to collate data and information about the laboratory's operations

Laboratory informatics is the specialised application of information technology aimed at optimising laboratory operations by the application of information technology to the handling of laboratory data and information. It encompasses four major multi-user systems: laboratory information management systems (LIMS), electronic laboratory notebooks (ELNs), laboratory execution systems (LES) and scientific data management systems (SDMS).

There is a very good reason why the use of a generic term such as 'laboratory informatics' is important: we need to get away from an application-centric approach and think of a fully integrated laboratory and its interaction with other company systems. The deployment of an ELN generally represents the final step in making a laboratory fully electronic, and hence raises the demand to connect up all laboratory systems. Being fully electronic and being fully integrated are two different things.

For most labs, being fully 'electronic' corresponds to an application-centric portfolio of 'systems' that were not necessarily designed to work together, and for which interoperability is hampered by the lack of standards and so has to be customised. A smart laboratory is an 'integrated' laboratory that is modular, based on standards, and is designed to facilitate connectivity, data sharing and collaboration.

Over the past few years, the informatics market has undergone two interesting developments; firstly, the previously separate LIMS and ELN sub-markets have started to overlap, causing a certain amount of confusion to the application-centric mind-set; secondly, mergers and acquisitions have reshaped the vendor line-up, specifically in the ELN field.

The origins of the LIMS market can be traced back several decades to the point where the increasing prevalence of computers in the laboratory, coupled with their increasing processing power, led enterprising scientists to develop simple, custom computerised workflow systems to operate in conjunction with data acquisition

and data processing. In the early 1980s, first-generation commercial LIMS started to appear, usually based on minicomputers, supporting sample and test management, and reporting of results.

A second generation of commercial LIMS started to appear in the late 1980s, typically taking advantage of relational databases to provide more sophisticated functionality. The development of client-server based systems represented the next (third) generation of commercial systems, taking advantage of the evolution of the personal computer. The fourth generation emerged as the internet and wireless connectivity developed, offering opportunities to extend the reach of LIMS beyond the confines of the laboratory.

As LIMS products were increasingly adopted by laboratories, three specific additional requirements gradually became apparent. Firstly, there was a need to transfer data from laboratory instruments directly to the LIMS, to avoid transcription errors; secondly, the need to manage the instrument data files from which data stored in the LIMS was derived; and thirdly, the need to handle unstructured data, graphical data, and to collate sample data. These requirements led to the development of scientific data management systems (SDMS) and electronic laboratory notebooks (ELNs). Functionally, the LIMS products have become increasingly sophisticated, to the point that the dividing line between LIMS and other informatics products has become less clear.

The initial evolution of the ELN market was centred on the provision of functionality to support small molecule chemistry

“

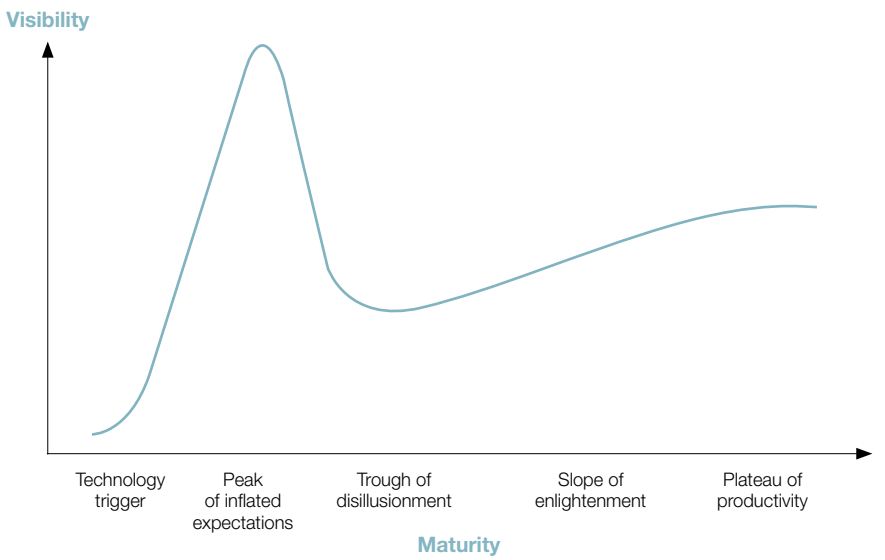
The ELN market has grown and developed rapidly over the past decade, it still exhibits some instability with a large number of vendors (there are more than 30 purveyors of products that purport to be an ELN) competing for market share. As a consequence, the market suffers from some degree of ‘hype’ (see Figure 4).

Just where ELNs sit on the Gartner Hype Cycle[1] is probably somewhere around the ‘Trough of Disillusionment’, although individual vendors may occupy positions either side of this point. The ‘Trough of Disillusionment’ can be considered as the turning point past the hype and when the focus is on delivering true benefit. Chemistry-based and generic ELNs are probably already beyond this point, as indeed are the majority of LIMS products.

Commercial ELNs have evolved from two approaches: discipline-specific; and generic. Generic software provides the architecture and tools to create and search content, and to work collaboratively in a way that satisfies the needs of almost any science-related industry. Discipline-specific ELNs are aimed at a particular market segment, such as chemistry, biology, or analytical. These systems are usually tailored to work with other discipline-specific software tools. Most of the commercial ELNs offer a combination of generic and discipline-specific functionality.

The initial evolution of the ELN market was centred on the provision of functionality to support small-molecule chemistry. Most of the experimental processes associated with synthetic chemistry are well established, reasonably consistent, and are well supported by desktop software tools. Integrating these functions into an ELN that can create, manage and store a full experimental record was a logical progression. As a consequence, chemistry-based ELNs are well established and exhibit a good deal of maturity. If there is segmentation in this part of the market, it is determined to some extent by the origins and scope of the available products.

FIG 4 The Gartner Hype Cycles



Technology trigger: The first phase of a Hype Cycle is the ‘technology trigger’ or breakthrough, product launch or other event that generates significant interest.

Peak of inflated expectations: In the next phase, a frenzy of publicity typically generates over-enthusiasm and unrealistic expectations. There may be some successful applications of a technology, but there are typically more failures.

Trough of disillusionment: Technologies enter the ‘trough of disillusionment’ because they fail to meet expectations and quickly become unfashionable. Consequently, the press usually abandons the topic and the technology.

Slope of enlightenment: Although the press may have stopped covering the technology, some businesses continue through the ‘slope of enlightenment’ and experiment to understand the benefits and practical application of the technology.

Plateau of productivity: A technology reaches the ‘plateau of productivity’ as its benefits become widely demonstrated and accepted. The technology becomes increasingly stable and evolves in second and third generations. The final height of the plateau varies according to whether the technology is broadly applicable or benefits only a niche market.

Some, for example, will be an enterprise-wide solution, others will focus on utility and personal productivity, while others will provide a generic ELN capability that integrates third-party software tools.

Biology, however, has presented a bigger challenge to the ELN vendors. The diverse and complex nature of biological processes and outcomes creates a need to capture not just the data, but also the interrelationships between the data. This, coupled with a diverse portfolio of biology-specific software tools, begs the question: do biologists just need a generic ELN that will integrate with their existing software tools, or do they need a complete suite of functionality that is embedded in the ELN? The issue for the biologists is whether there is a commercial ELN that addresses their specific and diverse requirements. Furthermore, for those companies that need to support chemists and biologists, the question is whether it is possible to find a single vendor solution that addresses the requirements of both disciplines, or whether to choose the best of breed for each discipline.

Within the past two or three years, another ELN domain has emerged, that of QA/QC and the regulatory world. A few vendors have concentrated specifically on this area, with products that are strongly aligned to laboratory workflows, following the step-by-step execution of SOPs or test methods.

The products are more structured than a 'conventional' ELN and in some respects appear to be functionally closer to a LIMS. This particular segment of the market has seen a number of LIMS vendors extending the functionality available in their LIMS products to embrace some of the more unstructured requirements associated with experimentation. It could be argued that such products may be better labelled as laboratory execution systems (LES) as they follow a very prescriptive approach applicable to those communities engaged in regulatory based testing.

To replace a paper notebook, all that could be required could be a simple authoring tool capable of generating a compound-document. However, additional capability will be needed for storing and searching documents, and for addressing workflow requirements. Some organisations have chosen to implement generic ELN functionality within the framework of their standard IT tools, such as Lotus Notes and SharePoint. In the academic community, blogging tools have been used to record experimental work and thus provide the basic features of an ELN, with a strong emphasis on sharing and collaboration and in the form of a laboratory journal.

The convergence in the informatics market is now confusing potential customers. The table, right, identifies the core differences in the major tools.

Initially, each of these tools addressed a well-defined, functional requirement, but the increasing level of sophistication of the underlying information technologies has made it easier to extend functionality in ways that mean that there is now considerable overlap between the different tools. At one stage it was considered unlikely that a single ELN could provide the necessary functionality to support chemistry, biology and analytical requirements. Those days are over, and this should make the task of finding a suitable ELN easier. But the extent of the overlap with LIMS, SDMS, and LES can generate confusion, and for someone looking to address laboratory information management requirements, the task seems to be more challenging.

ELN	Experiment-centric: an authoring tool that handles unstructured data and offers generic and specific functionality to support different scientific disciplines. Supports IP protection, knowledge re-use, productivity and collaboration.
LES	Procedure or experiment-centric: basically able to handle structured data and some unstructured data. Specifically designed to meet the requirements of the GxP environment. Simplifies repeated operations. Supports electronic SOPs.
LIMS	Sample-centric: primarily designed to handle structured data, and offers sample and test management, batch operations, and industry-specific workflows. Secure laboratory information hub. Supports compliance.
SDMS	Data-centric: handles data files from laboratory instruments, meta-data, documents, and the relationships between them.

What is a laboratory information management system (LIMS)?

A laboratory information management system (LIMS) provides the basic functions for sample and test management, and has become the standard tool for analytical and QC laboratories for registering samples, assigning tests, gathering and managing results, and issuing reports. Most LIMS now provide a more integrated solution to support workflows and processes customised to a range of industry-specific requirements.

The basic functions to be found in a LIMS are:

- The registration of samples and associated data, such as provenance, customer, due dates, etc.;
- The assignment of tests to the sample;
- Scheduling and tracking of the sample and tests;
- Recording the test procedure, equipment and materials used during testing;
- The review, approval, and aggregation of test results for the sample, including specification checking ; and
- The preparation and communication of customer reports.

The major business benefits of a LIMS are typically associated with more efficient workflows by eliminating errors due to manual data entry and transcription errors. This is achieved through interfacing laboratory instruments to the LIMS for two-way communication of sample IDs, worklists, and results, and by integration with other laboratory systems such as electronic laboratory notebooks (ELNs) and scientific data management systems (SDMS).

A LIMS also acts as a major repository of the records of analytical testing and can be a source of historical data associated with the organisation's products and production processes. In addition, the transactional nature of a LIMS enables a secondary record system to be maintained as an audit trail to track date, time, user – and, if necessary, what change was made within the system. This data may then be used to satisfy quality assurance requirements in terms of data integrity, and can also be used to generate a wide variety of management reports on the laboratory's performance.

A pre-requisite before implementing a LIMS, or indeed any major computerised system, is to map and optimise the laboratory processes that the LIMS will automate. The laboratory needs to understand the process and to identify any bottlenecks and their underlying causes.

Most laboratory processes have evolved over time to meet local laboratory requirements rather than being specifically designed to meet wider organisational requirements. Any LIMS implementation must simplify and streamline the process rather than automate an inefficient, paper-based status quo.

The commercial systems on the marketplace have become increasingly sophisticated over the years. The major challenge in choosing a LIMS is identifying how an out-of-the-box solution is aligned to the organisation's needs. Most systems are highly configurable and avoid the need for any custom code to be written to meet specific requirements.

What is a scientific data management system (SDMS)?

A scientific data management system (SDMS) is, in its basic form, a database application that manages electronic records generated by laboratory instruments. Typically, an SDMS will provide long-term data preservation, accessibility and retrieval. It is complementary to other laboratory informatics systems, such as LIMS and ELNs, in the sense that it can provide a common repository for experiment- and sample-related data files. In this way it provides a more consistent approach to managing laboratory data than local repositories, and off-line media (CDs, DVDs, tape, etc.)

The lines between a LIMS, ELN and an SDMS are at times blurred through the incorporation of additional features to complement the core functionality. An SDMS is a means of collecting data files from a wide range of laboratory instruments and storing them, along with metadata, in a uniform way in a database; in other words, it is a laboratory content management system. By adding workflow elements and providing facilities for the management and storage of other documents associated with laboratory operations (worksheets, SOPs, safety information, reports, PDFs, office documents, images, etc.), an SDMS can in practice evolve into a more comprehensive single informatics solution for some laboratories. However, an SDMS is essentially an 'event-driven' system that gathers data, which may limit some of its capabilities relative to the other informatics tools, and is therefore more frequently seen as a system that is complementary to a LIMS or an ELN.

Nevertheless, the principle on which the SDMS is based is that it aggregates records into a logical collection associated with a specific entity such as a programme, project, experiment, product, or sample, to provide a readily accessible collection of relevant

information. Embedded into an SDMS will also be the means to provide appropriate security of the records by means of access control, audit trail, authorisation, and change management.

What is a laboratory execution system (LES)?

A laboratory execution system (LES) sits somewhere between an ELN and a LIMS in terms of the functionality it delivers, but its existence is typically targeted at analytical service and quality control laboratories where high-volume workflows and regulatory compliance are primary business requirements. In a very basic sense, the underlying logical structure of an LES is almost identical to a LIMS, but the user interface is procedure-centric, rather than the usual sample-centric approach found in a LIMS. This allows a standard laboratory operating procedure (SOP) to be executed in an automated way, usually by interaction with laboratory instruments interfaced to the LES (where possible) in order to capture data without the need for transcription. Calculations on the captured data can be performed in the system, and thus the automated approach can eliminate two potential sources of error. The concept of a 'paperless lab' is a specific objective of the LES, eliminating the use of paper either for intermediate recording of data, or for longer term record keeping and archival purposes.

The LES is designed to adhere to laboratory workflows and provides a more repeatable and structured approach to quantitative testing procedures to help ensure compliance. The user interface usually takes the form of an electronic equivalent to the paper version of a laboratory standard operating procedure or worksheet. This type of interface is often referred as 'paper on glass', a term also used for a generic electronic laboratory notebook. Most LES applications can be readily configured to support alternative laboratory workflows in a way that relates closely to traditional paper-based processes. Worksheets can be converted to electronic forms; standard operating procedures can be configured with appropriate data entry points, with data checking; and unique workflows can be mapped to support repetitive and routine procedures.

As with other laboratory informatics systems, the underlying information technologies can extend an LES to a broader range of capabilities. For this reason, the LES can, in some cases, serve as an alternative to a LIMS, an ELN, or an SDMS. As with each of the major laboratory informatics tools, purchasing and implementation decisions require a thorough understanding of the laboratory's functional requirements. However, it is more likely that the LES will be seen as complementary to ERP and QM systems where high-throughput QA is an essential step in a business process.

What is an electronic laboratory notebook (ELN)?

In its simplest form, an electronic laboratory notebook can be considered to be a direct replacement for the paper lab notebook. In this instance, it can provide the generic functionality ('paper on glass') to support scientific documentation for patent evidence, cross-discipline collaboration, and general record keeping. However, the integration capabilities raise the possibility of a tighter coupling of other laboratory systems into the 'electronic laboratory notebook'. In other words, can the information that is currently printed from other laboratory systems, cut out and pasted into the paper lab notebook, be electronically entered or linked directly to the electronic laboratory notebook?

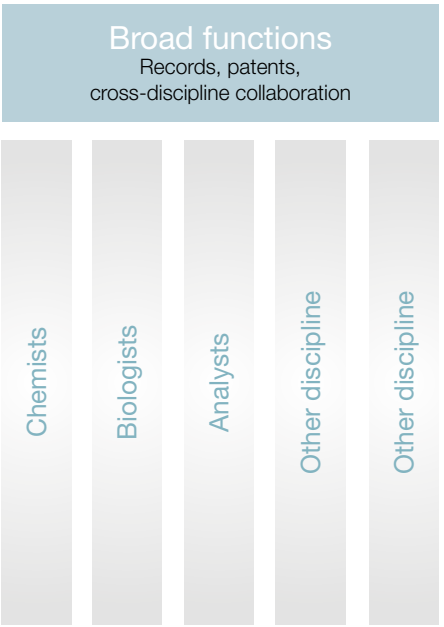
In the academic community, blogging tools have been used to record experimental work and thus provide the basic features of an ELN

"

For example, systems that provide chemical structure drawing, structure and sub-structure searching, and compound registration are an integral part of the chemistry laboratory's process, and therefore would be expected to become part of an electronic solution. Similarly, other scientific disciplines will have specific requirements consistent with their particular laboratory processes.

Figure 5 illustrates the relationship between 'broad' (generic) and 'deep' (specific) systems. In this context, the 'notebook' functionality (see Figure 1) is addressed by the 'broad' layer, whereas the discipline-specific functionality penetrates the 'interpreted/processed data' layer in Figure 1.

FIG 5 Broad vs. deep



From a patent perspective, the 'experimental layer' of Figure 1 is crucial as it captures what the scientist is thinking and doing, and therefore will provide the evidence of conception and reduction to practice of the 'invention'. In broader intellectual property (IP) terms, it is the 'experiment' layer that constitutes a record of the laboratory's work and as such contributes to the scientific knowledge repository.

For as long as this repository resides on paper, the ability to access, collaborate and share scientific knowledge is constrained. The implementation of an ELN therefore offers a significant opportunity to bring about greater efficiencies.

But a clearly defined understanding of the role that the ELN is going to play in a given organisation is absolutely essential at the start of an electronic laboratory notebook project. As discussed, an electronic laboratory notebook supports the 'experiments' layer, and also contains abstractions from the lower data levels (see Figure 1).

So the CENSA[2] definition of an electronic laboratory notebook as 'a system to create, store, retrieve and share fully electronic records in ways that meet all legal, regulatory, technical and scientific requirements' is all encompassing and can mean different things to different people.

Configuration versus customisation

The difference between customisation and configuration is very simply the difference between writing additional code and setting (configuring) in-built parameters in order to achieve some desired functionality. Customisation is generally considered a poor choice as it increases costs, complexity, and risk, and makes it more difficult and more expensive to upgrade software in the future. In a regulated environment, custom code will require extra validation steps. It may often be a symptom of bigger problems, including a mismatch with a company's requirements or a lack of project controls during implementation. Most laboratory informatics systems are designed to be configurable, and a major activity during implementation is to undertake the entire required configuration to meet functional requirements. Once configured, system upgrades will automatically carry through existing configuration.

An ELN can serve the organisation in three ways: it can take advantage of the capabilities of IT to improve the ability to acquire, manipulate, share and store data (productivity); it can facilitate communication and sharing in real-time across multi-disciplinary and multi-site teams (collaboration); it can provide a scientific knowledge repository that can be easily accessed to recover records of the laboratory's work (content/knowledge management).

The way in which laboratory notebooks are used is largely dictated by the United States' patent system which, unlike the rest of the world, is based on 'first to invent'. The need to be able to demonstrate who really was first to invent requires the laboratory notebook to be an authentic and trustworthy record that describes the concept and its reduction to practice, and for it to be signed by the author and corroborated by an impartial witness. There are two factors why the migration away from paper lab notebooks has taken so long: the reluctance of lawyers and patent attorneys to gamble on the legal acceptance of electronic records in patent interferences and patent litigation without any case law; and the lack of confidence in our ability to preserve electronic records over several decades.

An ELN system (like the bound laboratory notebook) has several roles:

A place to do science – a working environment;
A place to write up the experimental work;
A record of the work; and
A long-term preservation mechanism.

The wide range of commercial systems on the marketplace has become increasingly sophisticated over the years

"

One of the challenges to a successful ELN implementation is identifying exactly what role the ELN will play. The term 'electronic laboratory notebook' is inherently ambiguous. In most cases, the ELN is expected to do more than just replace the paper lab notebook. The paper lab notebook is a simple authoring tool, and any electronic authoring tool capable of generating a compound document will serve as a replacement.

For some companies this has proven to be the case. The combination of Microsoft Office, SharePoint services and a means of preserving documents (e.g. in PDF – portable document format) has proven to be an adequate replacement for paper. But if more functionality than this is needed – for example, integrating various chemistry or biology-centric functions, or other discipline-specific tools – then we are really talking about an electronic laboratory rather than an electronic laboratory notebook.

Chapter summary

The four major laboratory informatics systems serve different basic functional requirements, but convergence and increasingly sophisticated technologies are creating a good deal of overlap between the systems.

So when it comes to choosing the right solution, it's far better to start by defining an objective or describing the problem to be solved, rather than placing the initial focus on a 'solution'.

Just deciding 'we need an ELN' or 'we need a LIMS' should not be the starting point; it's far better to think about the big picture, i.e. the end-to-end business process that embraces the role of the laboratory, the specific workflows, the communication and collaboration requirements, and the integration requirements.

Once these requirements are defined, then the task of finding a solution is more straightforward. ■

KNOWLEDGE

Document management

This chapter considers how the smart laboratory contributes to the requirements of a knowledge eco-system, and the practical consequences of joined-up science. Knowledge management describes the processes that bring people and information together to address the acquisition, processing, storage, use, and re-use of knowledge to develop understanding and to create value

Scientists have been involved in the development of artificial intelligence (AI) for decades. The modern version of AI, which sought to create an artificial human brain, was launched in 1956 but had been fermenting many years before that, following the discovery that the brain was an electrical network of neurons that fired in pulses. Over the years, there were apparent breakthroughs that were followed by troughs of despair caused by hardware and software limitations. Among the highlights of this era was a Siri-like machine called ELIZA which, in 1966, could be asked natural language questions and provided voice-appropriate – albeit canned – answers.

The birth of the discipline of knowledge management (KM) was in the early 1990s. A tidal wave of KM consultants appeared, heralding the birth of this newer AI version and the emergence of supporting hardware and software. A few years later, several scholarly journals appeared as forums for advancing the understanding of the organisational, technical, human, and cognitive issues associated with the creation, capture, transfer, and use of knowledge in organisations.

Today the value of tapping into an easily accessible collection of information, such as a smart laboratory's assets, is appreciated more than in the past. The amount of data and information generated by instruments and scientists increases exponentially, and staff turnover is rising to the point where someone with seven years of service with the same employer is now considered an old-timer. Undocumented know-how and locations of information resources are now issues. Reinventing the wheel is becoming more commonplace – not a desirable occurrence, because the costs of drug development are ever-increasing and fewer blockbuster products are hitting the market.

Many software solutions offered to assist information management are specialised and fragmented. Often, disparate divisions of an organisation select their own local software solutions, and IT departments often dictate

requirements that restrict the scope of possible vendor solutions. Excluding small, single location labs, it is rare to see a smart laboratory where all associated information resides under one roof.

There are general solutions to support the processing of large data sets in a distributed computing environment. One of the best known is Hadoop, sponsored by the Apache Software Foundation. Hadoop makes it possible to run applications on systems with thousands of nodes, involving Petabytes of data. Its distributed file system facilitates rapid data transfer among nodes and allows the system to continue operating, uninterrupted, in case of a node failure. The Hadoop framework is used by major players including Google, Yahoo and IBM, largely for applications involving search engines and advertising.

Organisation is everything

When organising information one needs to decide what is important and what is not. Traditionally, in the paper notebook era, experiments, results, and comments were systematically entered to show diligence in pursuing a potential patent on an invention. Nothing could be removed; only subsequently noted or re-explained. Supporting data from instruments was retained with the notebook entries, and this practice led to the warehousing of innumerable papers as well as electronic records that might or might not be needed to support patent claims or meet regulatory requirements.

The volume of instrumental data today is much larger. Is it prudent to keep everything, or perhaps classify the data into two piles – one that directly supports a conclusion and another that is perhaps more generic? All electronic data suffers from aging, not unlike human aging. We'll talk about media and file format aging a little later, but we should also consider relevance aging. Should a particular spectral analysis file be kept or should the sample be re-run five years from now using updated equipment?

Information needs to be categorised into a small number of groups, preferably in a central location to facilitate retrieval.

Start with two piles and gradually split them appropriately. It is sensible to imagine how a researcher in the future would look for things, having no knowledge of past notations and conventions.

People like to use familiar visual signs to navigate. It's natural and usually results in finding what is needed plus additional, associated materials. Search engines may give more precise results but may omit important things that are part of the navigation journey. Scientists appreciate the role of serendipity in drug discovery.

Retention schedules

Not everything can be kept for ever – but how long is sensible? There is some consensus that information supporting a patent should be retained for the life of the patent, plus several years before and after to cover eventualities. Most pharmaceutical companies have settled on a 40- to 65-year retention for intellectual property. Records to support regulatory compliance sometimes need to be retained for as long as 25 years. At the end of their retention period, records should be evaluated for their disposition. Should they be destroyed, or perhaps kept for a few more years? Scheduled examinations of records have a bonus of providing information that could be applied to current issues. Looking through the supposed 'rubbish' can be a very good thing.

There are at least two good reasons for retention schedules. First, there is the smoking gun. In the event of legal or regulatory investigations and/or audits, there's bound to be information that is erroneous, that conflicts with established facts, or serves no particular purpose. Observations and comments that arae taken out of context can also be misleading. This is not a licence to cook the books; the aim is to throw out the junk and items that have no real contribution to the organisation. It's better to identify what needs to be retained before any of these issues occur. The non-records should be destroyed as quickly as possible and the declared records evaluated after a pre-prescribed time (retention period). Keeping non-records and records past their retention dates costs money. The cost of hardware associated with information storage continues to decrease but the amount of labour needed to support large collections has increased sharply.

People like to use familiar visual signs to navigate. It's natural and usually results in finding what you want plus additional associated materials

"

Authentication

How can records that are created within an organisation be authenticated? They don't all need to be notarised, but it would be nice if there was an easy way to come close to this. So here are the concepts to use. Appoint a designated records manager who will have full control of the records. People have been doing this for years with paper records – it works. The custodian authenticates the author and maintains a chain of custody if the record is moved. Copies can be made and distributed, but the 'original' is always in the vault. It's pretty much the same with electronic records: the documents are stored on a server where users can view them or make copies. The official, 'original' record stays in its slot. Chain of custody is maintained when the record is migrated to another location or is converted into other formats.

Long-term archiving: paper and microfilm records

There is a general perception that records will be easy to find, retrieve, and view in the distant future. Paper and microfilm records that are stored in a clean, temperature- and humidity-controlled environment could be readable for more than 100 years. However, finding and retrieving them requires some strategic planning. At the very least, they should be organised by year. Additional sub-categories or folders can be added to facilitate retrieval. The ideal solution involves the assignment of a unique identifier to each record; the identifier containing or linking to relevant metadata to aid in searching. For large collections, this information should be stored in a database. A plan must be developed to migrate this information from its existing hardware and software, after it becomes obsolete, to newer systems.

Long-term archiving: electronic records

We are all aware of the extremely short half-life of computer hardware and software. The software authoring tools in use today will blink out of existence and be replaced by tools that have more capabilities or are compatible with current operating systems. One can only speculate regarding the hardware and data storage media we will be using in the future. There will probably be no practical Rosetta Stone to help translate codes used in legacy software. Maintaining authenticity and minimising data corruption needs to be addressed.

There have been attempts to maintain a museum of hardware and software that could help in viewing legacy records. These mostly failed, most notably an effort by the National Aeronautics and Space Administration (NASA).

NASA lost many of its electronic records from the early 1960s and then took steps to ensure that it would not happen again.

This resulted in the 2001 launch of the Open Archival Information System (OAIS) reference model, sponsored by a global consortium of space exploration agencies concerned with data preservation.

Other global consortia have come together to develop preservation strategies. The International Research on Permanent Authentic Records in Electronic Systems (InterPARES) aims at 'developing the knowledge essential to the long-term preservation of authentic records created and/or maintained in digital form and providing the basis for standards, policies, strategies and plans of action capable of ensuring the longevity of such material and the ability of its users to trust its authenticity.'

Finally, Australia's Victorian Electronic Records Strategy (VERS) provides a framework within which to capture and archive electronic records in a long-term format that is not dependent on particular hardware or software.

The concepts that these global data initiatives use for long-term preservation are the same. First, capture the content and metadata, then protect them with an immutable file format that preserves the text, images, charts and tables and renders them readable in the way the authors intended. The emerging standard for this purpose is PDF/A, an ISO-standardised version of the portable document format (PDF). Finally, the immutable file is further protected from tampering by digital encryption.

Electronic storage media is a moving target. It is quite unlikely that media being employed today will be used beyond the next 20 years. The storage of electronic information on magnetic tape, pioneered by IBM in the 1970s, is not only the storage method of choice today, but its usage is increasing.

Tape is far cheaper and more reliable than any other medium used for archiving data. This does not mean that records from a 20-year-old tape can be retrieved readily unless a compatible drive, which could retrieve its content, has been saved in a museum.

To keep electronic records for more than 10 years, a migration strategy needs to be developed and implemented now, before the museum closes.

Chapter summary

The best approach to organising information is to decide what is important to keep and what is not. How would a researcher in the future look for things, having no knowledge of past notations and conventions? There are at least two good reasons for applying retention schedules. In the event of legal or regulatory motivated investigations, and/or audits, there's bound to be information that is erroneous, conflicts with established facts, or serves no particular purpose.

If there is a risk that observations and comments can be taken out of context, items that have no real contribution to the organisation's business should be thrown out. Records that are past their retention dates should also be discarded to avoid maintenance costs.

The cost of hardware associated with information storage continues to decrease, but the amount of labour needed to support large collections has increased sharply. A records manager should be designated and given full control of the records.

The basic guidelines are as follows:

- Understand the legal implications of electronic records;
- Establish a file plan;
- Establish an electronic records preservation file plan;
- Establish an electronic records manager or management team;
- Establish and communicate policies;
- Avoid point solutions; and
- Don't keep electronic records forever. ■

FURTHER AFIELD

Beyond the laboratory

This chapter considers who cares about how smart the laboratory is, and why? It also looks at the broader business requirements and their impact on the laboratory, with an emphasis on productivity and business efficiency, integration with manufacturing and business systems, patent evidence creation, regulatory compliance, and data integrity and authenticity

There is a quote that states: 'A couple of days in the laboratory can save a couple of hours in the library'. This sentiment once typified the attitude of a lot of scientists, but those days are over. It may not have reached the point where a couple of hours on the computer can save a couple of days in the laboratory, but it's heading that way. The laboratory is part of a business process and, as such, it is subject to the same productivity and efficiency targets that apply to other parts of the business.

Most laboratory informatics projects focus on the return on investment, typically quantified by streamlining data input through the elimination of bottlenecks, by interfacing systems, and by removing manual processes involving paper. Most projects will also specify long-term gains through establishing a knowledge repository, but this is where quantitation becomes difficult. It's not unusual in the early days of an informatics deployment, for example, for a simple search to uncover prior work that can save reinventing the wheel. But, as time goes on, it becomes the norm to check before starting an experiment. There is, however, another side to exploiting the knowledge base – which we're only just starting to come to terms with. The need to delve deeper into the knowledge base to visualise and interpret relationships and correlations is growing. 'Big data' is the popular term being assigned to the data problem, as it applies to all walks of life. This puts the emphasis on ensuring that we have confidence in the integrity, authenticity, and reliability of the data going in, and that the appropriate tools are available to search, analyse, visualise, and interpret the information coming out.

These requirements are driven by the requirements of the laboratory's customers for robust, reliable, and meaningful scientific information and data that is delivered in a timely and cost-efficient way. The time-honoured principles of the scientific method provide the basis for the integrity, authenticity, and reliability of scientific data, but those principles need to be reinforced in the context of regulatory compliance and patent evidence creation.

Productivity/business efficiencey

The basic objective in deploying laboratory informatics systems is to improve laboratory productivity and business efficiency. To maximise the benefits, it is important to consider the wider laboratory and business processes

that may be affected by the new system. It is easy to fall into the trap of just 'computerising' an existing laboratory function, rather than looking at the potential benefits of re-engineering a business process. The use of tools such as 6-Sigma or Lean can help considerably. Nevertheless, it is prudent to be careful with the use of these tools, depending on the nature of the lab. For example, high-throughput, routine-testing laboratories, which basically follow standard operating procedures, are more receptive to process improvement. Discovery/research laboratories however, which are less structured and are dependent on more diverse and uncontrolled processes, are less likely to benefit from formal process re-engineering.

Productivity and business efficiency are usually measured in financial terms, although this may be translated into time-savings or, in some cases, the numbers of tests, samples, experiments completed. It is necessary, therefore, to be able to quote 'before and after' figures for any deployment project. Establishing a baseline metric is an important early step in the project.

The tools can facilitate improvement through well thought-out deployment, but also offer the capability to monitor and improve processes.

Costs/return on investment

Any organisation considering the implementation of a new informatics or automation system will want to investigate the return on investment (ROI), or cost/benefit. This is usually extremely difficult, since many of the projected benefits will be based on a certain amount of speculation and faith. However, there are some important points to consider in building the cost/benefit case. The costs associated with managing paper-based processes (e.g. notebooks, worksheets, etc.) through their full lifecycle in the lab are not always fully visible or understood.

Apart from the material costs, and the costs of the archive process, there is a hidden cost – and the time taken in writing by hand, cutting, pasting, transcribing, and generally manipulating paper, as well as approval and witnessing processes, all contribute to this hidden cost. It is normal in building the cost/benefit equation to look at how much of a scientist's time is spent managing the paper-based processes, and to use this as a basis for potential time-savings with an electronic solution (see Figure 6). Although the start-up costs are high for an electronic solution, the incremental cost of adding new users and increasing storage space is modest.

ROI tends to focus on the short term: how soon can one get a return on the money invested in deploying a new system? But the true value of the system may be in the long term and, therefore, far more difficult to measure as the value will be determined by behavioural changes. There is a growing body of evidence being presented at conferences on electronic laboratory notebooks (ELNs) by numerous companies that have implemented them, showing that the short-term time savings associated with the electronic solution are significant. These organisations also list a number of other non-quantifiable, long-term benefits such as:

- Scientists spending more time in the laboratory;
- It is easier to find information in a searchable archive;
- It is easier to share information;
- Increased efficiency through the elimination of paper;
- A reduced need to repeat experiments (knowingly or unknowingly);
- Improved data quality;
- A smooth transition when people leave the company; and
- Online use in meetings.

Building a good business case requires a thorough and systematic approach to understanding current limitations as well as future requirements for the business

"

Regulatory compliance

The early research phases in the pharmaceutical industry comprises the testing of large numbers of chemicals to see if any of them have potential as a new drug. Only the best will go on to more extensive testing. There has been a 'consensus' that regulatory work does not start until the chemical has been chosen. Then adherence to GMP[3] (good manufacturing practice) and GLP[4] (good laboratory practice) starts, and the IT systems need to be in compliance with the local requirements for IT systems. In the US, this is 21 CFR Part 11[5] and in the EU it is GMP Annex 11.[6] While this may be at least partially correct, the fact is that the data, and of course the IT systems that hold the data, need to be under control for another business reason: patents.

The US patent system is based on 'First to Invent', and that means it must be possible to prove the date of the invention. Traditionally, this has been done using bound paper notebooks, where the entries have been dated and signed, and co-signed by a witness. Paper notebooks can be admitted as evidence if they can be demonstrated to be relevant. Electronic records are equally relevant, as the judgment is made on the evidence – not on the medium that holds the evidence. One important factor is the data integrity, which must be possible to prove in court if necessary. Bound paper logbooks are still being used to a large extent, as most legal advisors don't feel comfortable with electronic data. It may be smart to talk to patent lawyers before starting to create electronic lab data.

How can we prove that the IT system is good enough?

The answer is, of course, validation. Actually, validation of processes is nothing new; that has been a part of the GMP and GLP regimes since they were introduced. An IT system is a part of the process and must therefore be validated as well.

The industry asked the US Food and Drug Administration (FDA) how it would handle electronic signatures, and accordingly 21 CFR Part 11 saw the light in 1997. The surprise was that most of the two-page document was about electronic data, and only a little about signatures.

This, however, does make sense. How can scientists use an E-signature if they are not sure that the data is (and will be) valid? They can't; they need to have control of your data before they can sign it electronically. The EU also came up with an equivalent to 21 CFR Part 11, namely the EU GMP Annex 11. This was revised in 2011 but does not improve on the first version. But a really good document covering electronic data and signatures is yet another document numbered 11, the PIC/S PI 011.[7] This is a 50-page document with the same requirements as Part 11, but it includes also a lot of explanations. PIC/S is the organisation for European pharma inspectors. They do stress that this document is not a regulatory requirement, only an explanation to the inspectors on how to handle IT systems. How that cannot be a requirement document, is hard to understand, however. The main difference between Part 11 and Annex 11 is that the latter also includes risks. IT validation shall be based on risks; high-risk systems need more validation than low-risk systems.[8]

This follows the same line of thought that the FDA started in the early 2000s: know your processes, and base the work on the risks they encompass.

How do we validate our IT system?

The best answers are in a guidebook called GAMP5[9]. GAMP also has a few more detailed sub-books.

What is written in this guide to a smart laboratory is of course just an overview. Please see GAMP5 for more details.

The GAMP5 way of validating the IT systems is as follows:

- Risk management to decide how important the system is in the process;
- Categories of software to decide what needs to be done;
- Combination of risks and categories to decide what to do for this system; and
- Testing guide for how to test the system.

Risk management

Identify regulated E-records and E-signatures:

- Is the record required for regulatory purposes? Is it used electronically? Is a signature required by GMP/GLP/GCP?

Assess the impact of E-records:

- The classification of potential impact on patient safety and/or product quality: is it high/medium/low?

Assess the risks of E-records:

- The impact and likelihood/probability of problems being detected/happening: is it high/medium/low?

Implement controls to manage risks:

- Modify processes, modify the system design; apply technical or procedural controls.

Monitor effectiveness of controls:

- Verify effectiveness, consider if unrecognised hazards are present; assess whether the estimated risk is different and/or the original assessment is still valid.

GAMP5 software categories

Category 1: Infrastructure software

- Definition: layered software (i.e. upon which applications are built). Software used to manage the operating environment.
- Example: operating systems, database engines, statistical packages, programming languages.
- Validation: record version and service pack. Verify correct installation.

Category 2: This category is no longer in use

Category 3: Non-configured products

- Definition: off-the-shelf solutions that either cannot be configured or that use default configuration. Run-time parameters may be entered and stored, but the software cannot be configured to suit individual business processes.
- Example: firmware-based applications, COTS, instruments.
- Validation: the package itself. Record version and configurations, verify operations against user requirements. Consider auditing the vendor. Risk-based tests of application: test macros, parameters, and data integrity.

Category 4: Configured products

- Definition: software, often very complex, that can be configured by the user to meet the specific needs of the business process. Software code is not altered.
- Example: LIMS/SCADA/MES/MRP/EDMS/ clinical trials, spreadsheets and many others (See GAMP5).
- Validation: life cycle approach. Risk-based approach to supplier assessment and other testing. Record version and configuration, verify operation against user requirements. Make sure SOPs are in place for maintaining compliance and fitness for\ intended use, as well as for managing data.

Category 5: Custom applications

- Definition: software, custom-designed and

coded to suit the business processes.
- Example: it varies, but includes all internally/externally developed IT applications, custom firmware and spreadsheet macros. Parts of Category 4 systems may be in this category.
- Validation: the same as Category 4, plus more rigorous supplier assessment/audit, full life cycle documentation, design and source code review.

System validation

The validation itself usually needs to be divided into more manageable pieces. One way is to use the '4Q method'. This comprises development qualification (DQ), installation qualification (IQ), operation qualification (OQ), and performance/process qualification (PQ). What the chosen manageable pieces or phases are is up to the individual, but it must be described in the validation plan. This defines the phases, the input and output of the phases, and which documents will be created during the phases. Typically, this will be a phase plan including test plans, the testing itself and the test documentation, and the phase report.

DQ – development qualification

This includes writing the user requirements specification, choosing the system, auditing the supplier if the risk assessment says that this is needed, and implementing the system.

IQ – Installation qualification

Installing the system is usually just a matter of following the description from the supplier. A brief test will show that the system is up and running.

There is a growing level of interest in how consumer technologies can enhance the user experience of working with laboratory informatics tools

"

OQ – Operation qualification

This may be defined differently in different organisations, but a common definition is to test each function separately. This is often what the supplier already has done, so the user may not have to do this if documentation and/or a supplier audit has shown that this has been done.

PQ – Performance or process qualification

The PQ is also defined differently in different organisations. Basically, the PQ is testing that the implemented system is according to business processes. This includes indirectly testing that the separate functions work as intended. This may also be called the system testing.

If the supplier's OQ testing is unavailable, more of the functions may have to be included in testing. It is perfectly fine to combine the OQ and PQ into one combined phase.

It is important to qualify or validate all the functions needed for your workflows.

A thorough description of IT validation can be found in the book, *International IT Regulations and Compliance*. This book also has chapters on LIMS and instrument systems, and the tips there are useful to read and follow in order to get a really smart laboratory with the information required.

But validation is never done. It's important to prove that the system is still validated, even after changes in and around the system. Having appropriate procedures to explain how to keep the validated state, and documentation to prove that procedures have been followed, are a must.

These procedures need to cover whatever is appropriate, including:
- Error handling, including corrective action and preventive action;
- Change management;
- Validation/qualification of changes;
- Backup and recovery;
- Configuration management;
- Disaster recovery and business continuity;
- E-signatures;
- Environmental conditions;
- Risk assessment and management;
- Security and user access;
- Service level agreements;
- System description;
- Training;
- Validation and qualification;
- Supplier audit;
- Daily use;
- Implementation of data in the system;
- Qualification/validation of implemented data in the system; and
- Data transfer between systems.

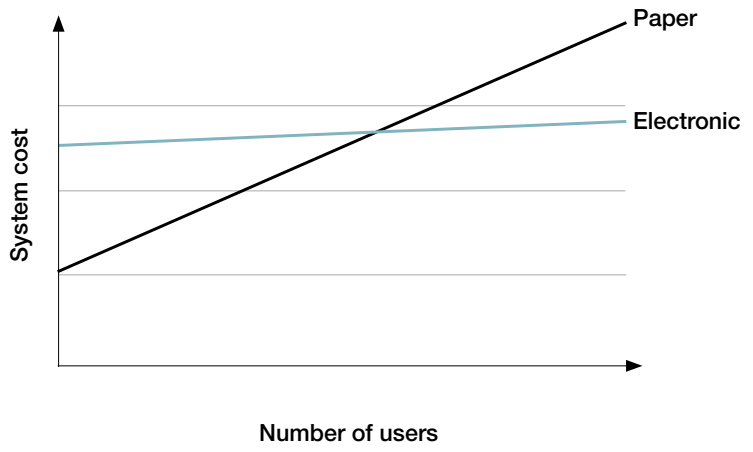
Some of these SOPs will be generic in the organisation, and some will be system specific.

Validation is a never-ending job, but with a validated system the user can be sure that the system works as intended and that the data is secured inside the system. That means that the user can prove beyond doubt that the data was entered on a given date and that the system will show that data has been corrected later.

Patent-related issues

The US patent system is based on 'First to Invent' and, in order to help determine who was first to invent, most companies engaged in scientific research create and preserve evidence that they can use to defend their patents at a future date. Traditionally, this evidence has been in the form of the bound paper laboratory notebook. In a patent dispute, any inventor is assumed to have an interest in the outcome of the case, so their testimony must

FIG 6 System costs of paper notebooks and ELNs



be corroborated. Most organisations require these notebooks to be signed by the author ('I have directed and/or performed this work and adopt it as my own') and also by an impartial witness ('I have read and understood this work').[10, 11]

Evidence in US patent interferences is subject to the Federal Rules of Evidence. There are a number of important hurdles that need to be overcome, in particular the Hearsay Rule (by definition, if the author cannot be present, then the evidence is hearsay) and the Business Records Exception.

The Business Records Exception is an exception to the hearsay rule, which allows business records such as a laboratory notebook to be admitted as evidence if they can be demonstrated to be relevant, reliable and authentic. The following criteria must be met:

- Records must be kept in the ordinary course of business (e.g. a laboratory notebook);
- The particular record at issue must be one that is regularly kept (e.g. a laboratory notebook page);
- The record must be made by or from by a knowledgeable source (e.g. trained scientists);
- The record must be made contemporaneously (e.g. at the time of the experiment); and
- The record must be accompanied by testimony by a custodian (e.g. company records manager).

Any doubt about the admissibility of electronic records was largely removed by this statement from the Official Gazette (10 March 1998 [12]:

'Admissibility of electronic records in interferences: Pursuant to 37 CFR 1.671, electronic records are admissible as evidence in interferences before the Board of Patent Appeals and Interferences to the same extent that electronic records are admissible under the Federal Rules of Evidence. The weight to be given any particular record necessarily must be determined on a case-by-case basis.'

In terms of admissibility, paper and electronic records are therefore equivalent. The judgment is made on the evidence, not the medium in which it is presented. However, it is important to understand the factors that impact upon the authenticity of electronic records and that in the adversarial nature of the courtroom, the opposing side may attempt to discredit the record, the record-keeping system, and the record-keeping process. The integrity of the system and the process used to create and preserve records are therefore paramount.

Many organisations still require their scientists to keep bound laboratory notebooks. This is because there isn't the case law or other experience for most legal advisors to feel as comfortable with electronic records as they are with paper. The issue is not one of admissibility, but of the weight that the record will have in court. Unfortunately, we are unlikely to see a suitable body of case law for many years.

The high-stakes nature of the problem, lack of experience, and long-term accessibility concerns have caused a number of organisations to adopt a hybrid solution, using an electronic lab notebook (ELN) front-end tool to create records, and then preserving the resulting records on paper. This gives the benefits of paper records (for the lawyers) while providing the scientists with the benefit of new tools. A fully electronic system will require scientists to sign documents electronically, and the resulting record to be preserved electronically.

Using multiple systems for patent evidence creation and preservation can expose an organisation to increased risk, due to the need to maintain the integrity of each system, and the consistency of the content between them. Similarly, the use of generic systems for such a task can increase discovery concerns and also increase the likelihood of problems. Further guidance should be sought from records management personnel and legal advisors within the organisation, in order to determine policy.

A recommended approach to help uncover and resolve legal/patent concerns is to work with the company's lawyers and patent attorneys to simulate the presentation of ELN evidence in the courtroom, and then work back to the creation of that evidence in the laboratory.

The America Invents Act – implications

Patent-reform legislation, in the form of the Leahy-Smith America Invents Act 2011, changed the US system from First to Invent to First to File in March 2013. It is very tempting to view this change as an opportunity to relax some of the procedural requirements of ELNs used in research laboratories.

However, there are clauses in the Act that would suggest it's wise not to make such an assumption. It is likely that patent interferences and interfering patent actions will continue for many years for patents and applications filed after March 2013. [13]

There are specific circumstances described in the America Invents Act that, for example, require proof of inventive activities to remove prior art for joint research activities, or to preserve the right to an interference if the application contains, or contained at any time, a claim to an invention filed before March 2013. Until the act becomes effective, and there is clarification about the implications of the new legislation, there is no reason to change in-house procedures for keeping laboratory notebooks, or for vendors to revise the procedures and workflows in their ELN products. The more immediate concerns are:

- There is a loophole that will allow people to prosecute a patent under the old First to Invent rules for many years to come. First to File isn't dead even after 16 March 2013 – there are some changes that mean proof of inventive activities will be especially important for joint research activities. The retention of other documentation related to joint research projects may need to improve; and
- Derivation proceedings will also require proof of inventorship.

To add further uncertainty, there's always a chance (or indeed probability) that things are going to end up in the US Supreme Court to examine the constitutional implications of a move away from First to Invent. So it does

appear that the new Act makes legally robust, signed, and witnessed records of inventive activities (generally in the form of lab notebooks) even more critical. With a move to 'First to File' there's the additional pressure of getting to the Patent Office quickly, which means it is necessary to start paying attention to the patent filing process, which has historically not been under much time pressure.

Data integrity, authenticity and management

Whenever electronic records are used within the framework of legal or regulatory compliance, data integrity and data authenticity are fundamental requirements of the computer systems used to create, manipulate, store and transmit those records. These requirements may also apply to in-house intellectual property (IP) protection requirements. It will therefore be necessary for a laboratory informatics implementation project to very carefully consider the specific requirements of their organisation in this area. [14]

The characteristics of trustworthy electronic records are:

- Reliability – the content must be trusted as accurate;
- Authenticity – records must be proven to be what they purport to be, and were created and transmitted by the person who purports to have created and transmitted them;
- Integrity – must be complete and unaltered, physically and logically intact; and
- Usability – must be easily located, retrieved presented and interpreted.

Data integrity, in a general sense, means that data cannot be created, changed, or deleted without authorisation. Put simply, data integrity is the assurance that data is consistent, correct and accessible. Data integrity can be compromised in a number of ways – human error during data entry, errors that occur when data is transmitted from one system to another, software bugs or viruses, hardware malfunctions, and natural disasters.

There are many ways to minimise these threats to data integrity including backing up data regularly, controlling access to data via security mechanisms, designing user interfaces that prevent the input of invalid data, and using error detection and correction software when transmitting data.

Data authenticity is the term used to reinforce the integrity of electronic data by authenticating authorship by means of electronic signatures and time stamping.

Generally speaking, electronic signatures are considered admissible in evidence to ensure the integrity and authenticity of electronic records. An electronic signature is a generic term used to indicate 'an electronic sound, symbol or process attached to or logically associated with a record, and executed or adopted by a person with the intent to sign the record.'

A digital signature is a specific sub-set of an electronic signature that uses a cryptographic technique to confirm the identity of the author, based on a username and password and the time at which the record was signed. The requirements for an informatics project will be somewhat dependent on the nature of the organisation's business and internal requirements, but security, access control and electronic signatures are factors that must be given appropriate consideration.

There are a number of ways to ensure data integrity and authenticity. The first is to develop clear, written policies and procedures of what is expected when work is carried out in any laboratory; the integrity of the data generated in the laboratory is paramount and must not be compromised. This is the 'quality' aspect of the quality management system (QMS) that must be followed.

There is the parallel need to provide initial and ongoing training in this area. The training should start when somebody new joins the laboratory, and should continue as part of the individual's ongoing training over the course of their career with the laboratory.

To help train staff, we need to know the basics of laboratory data integrity.

The main criteria are listed below:

- Attributable – who acquired the data or performed an action, and when?
- Legible – can you read the data and any laboratory notebook entries?
- Contemporaneous – was it documented at the time of the activity?
- Original – is it a written printout or observation or a certified copy thereof?
- Accurate – no errors or editing without documented amendments;
- Complete – all data including any repeat or reanalysis performed on the sample;
- Consistent – do all elements of the chromatographic analysis, such as the sequence of events, follow on and are they date- or time-stamped in expected sequence?
- Enduring – they must not be recorded on the back of envelopes, cigarette packets, or the sleeves of a laboratory coat but in laboratory note books and/or electronically by the chromatography data system and LIMS; and
- Available – for review and audit or inspection over the lifetime of the record.

It is important that laboratory staff understand these criteria and apply them in their respective analytical methods regardless of working on paper, hybrid systems or fully electronic systems. To support the human work, we also need to provide automation in the form of integrated laboratory instrumentation with data handling systems and laboratory information management systems (LIMS) as necessary. In any laboratory, this integration needs to include effective audit trails to help maintain data integrity and monitor changes to data.

Supervisors and quality personnel need to monitor these audit trails to assess the quality of data being produced in a laboratory – if necessary a key performance indicator (KPI) or measurable metric could be produced.

There is a growing level of interest in how consumer technologies can enhance the user experience of working with laboratory informatics tools

"

Chapter summary

From a broader business perspective, the introduction of computerised tools for managing laboratory information comes at a perceived higher cost, and challenges the user to consider very carefully the consequences of moving from a paper-based existence to one based on technology.

The return on investment equation is critical in obtaining the initial go-ahead for an informatics project, but the transition to digital from paper represents a major upheaval to long-established and well-understood information management processes.

Computer systems used in regulated environments need to be validated; the user needs to be confident that computerised systems can deliver productivity benefits, and data integrity and data authenticity can be guaranteed in a digital world.

Lawyers and patent attorneys need to be confident electronic lab notebooks can be presented as evidence in patent submissions, interferences and litigation. ■

PRACTICAL CONSIDERATIONS

Specifying and building the smart laboratory

This chapter looks at how to build a smart laboratory; what approaches to take; and how to deal with potential problems. Inevitably, becoming 'smart' takes time, not only due to the level of investment required, but also because of the impact of change and the need to consider legacy requirements. The rate of change in computer technologies is far greater than in the laboratory and in business, and this unavoidably means that the computing experience in the laboratory will lag behind the consumer experience. Additionally, the constraints of IP, regulatory and legal compliance do not lend themselves to risk-taking when deploying new technologies. New laboratory informatics projects demand a carefully managed and risk-averse approach

Functional/user requirements

Gathering user or functional requirements is one of the key tasks, usually assigned to the project team, to provide a specification against which potential solutions can be evaluated. The task involves uncovering and understanding user-needs, distinguishing them from 'wants' and 'nice to haves', and aggregating the needs into a requirements specification. In this context, reference to 'users' includes not just end-users of the proposed system, but anyone who will interact with the system, or be involved with inputs or outputs to the system. In order to do this, various methods may be used to gather needs and to prioritise them.

The requirements may include, but are not limited to:

- General business requirements;
- User/functional requirements;
- IT requirements;
- Interface requirements;
- Regulatory issues;
- Data management requirements;
- Error handling;
- Reporting requirements; and
- Performance requirements.

The criteria that define required performance may include:

- Access control and security;
- Look and feel;
- Robustness;
- Scalability;
- Ease of use;
- Technical performance/response times; and
- Technical support.

All of these requirements are normally collated into a request for proposal (RFP) that will be submitted to potential vendors. The RFP should also provide more general information, including an introductory description of the organisation and the major objectives of the project, as well as diagrams showing relevant workflows. The RFP may be preceded by a request for information (RFI) – a means of gathering information about a potential vendor's products and services, which may be used to fine-tune a final list of vendors to whom the RFP may be submitted.

Unfortunately, users are notoriously bad at stating what they need. Most systems are specified or designed by a team or committee and the team/committee members tend to be

volunteers who are committed to the concept of the system, enthused about the improvements it can bring, and are able to envision the potential. Unfortunately, the committee process can create complex systems and reflect compromises, and it is often the case that most problems come from people who don't volunteer for the committee! By definition, the members of the team are more committed to the success of the project than those who are not directly involved. In their deliberations, project teams often develop a concept of a solution that is much more sophisticated than might be needed or, indeed, is economically justifiable.

Typically the requirement-gathering phase involves harvesting needs, wants and ideas from the potential user community, and then engaging in a prioritisation exercise to reduce the list to a specific set of requirements that form the basis of a request for proposal (RPF) to be presented to vendors.

It is important that the business requirements are fully clarified first of all; this ensures that the scope of the project is defined and can therefore help exclude some of the more exotic 'needs' that might arise. Any single item on the requirements list should justify itself not only financially, but also in terms of its usefulness and ease of use.

Anecdotal experience suggests that some requirements specifications could be shrunk by between 25 and 50 per cent by the removal of 'wish list' items – bringing cost-savings and lower cost of ownership, as well as easier user adoption. It is important for the project team and sponsors to be able to define what business problem the electronic laboratory notebook (ELN) will solve, and to ensure that user requirements are kept simple and are focused on solving the problem.

The formal RFI and RFP approaches can only go so far, and it is essential that candidate systems be demonstrated and assessed with some preliminary configuration to establish and evaluate not only whether the system meets the functional requirements, but also whether it provides an acceptable user experience. In some respects, it makes sense to consider functional requirements and user requirements as separate criteria.

Business case development and project management

Building a good business case requires a thorough and systematic approach to understanding current limitations as well as future requirements for the business. It is important to see laboratory informatics as a component in a laboratory ecosystem (technology, processes and people), rather than 'just another laboratory application'. The following points should all be considered in formulating the case for a new informatics system.

Why do we need a new system?

- What is the problem that needs to be solved?
- Is there any quantitative data that illustrates the problem?
- Which laboratory areas will be involved in the project?
- Who makes the go/no-go decision?
- What are the issues relating to IP (internal/legal/patent)? and
- Are there any regulatory compliance requirements?

Clarify why the organisation thinks it needs a new system. This is best achieved by developing a problem statement that quantifies a specific problem, or set of problems, about the laboratory's productivity and/or knowledge management performance.

The scope and scale of the problem (and hence, the solution) should be identified. The key decision-makers/budget-holders should also be identified, plus any other interested party who may have influence over a go/no-go decision. It is important to know what business level constraints may apply in terms of internal, legal or regulatory compliance.

Building a good business case requires a thorough and systematic approach to understanding current limitations as well as future requirements for the business

"

Laboratory/company background

- Use organisation charts to clarify roles and responsibilities and organisational relationships;
- Identify the nature and scientific disciplines of the laboratory work and how they relate to each other; and
- Establish whether outsourced agencies (contract labs) are involved?

Establish the way in which the laboratory is organised, the nature of the work it undertakes and how it relates to internal and external organisations with whom it collaborates.

Current laboratory processes and systems

- Which laboratory systems are already in use? (Are there SOPs?)
- Which data acquisition systems are already in use?
- Which teamwork/collaboration systems are already in use?
- Which document management systems are already in use?
- Who is responsible for the management and support of these systems?
- Is there a (electronic) records management policy? and
- Are there any specific policies an restraints relating to the introduction of ITsystems?

Establish how the laboratory is currently working, paying specific attention to the use and effectiveness of manual systems such as worksheets, paper lab notebooks, and data management. Also identify major 'electronic' systems used for the acquisition, processing and management of data, and ask what happens to this data – where is it stored and for how long? Is it communicated or transferred elsewhere – if so, how? Is it backed up and/or archived? Can it be found?

Is laboratory data the responsibility of the laboratory, or does IT have any involvement? What level of involvement does IT have in the purchase and implementation of laboratory systems?

Future laboratory processes and systems

- Based on interviews with laboratory managers and laboratory staff, a model should be developed to illustrate the major relationships between laboratory data and information;
- Construct data workflow and laboratory process diagrams;
- Identify any conflicts in nomenclature and establish an agreed taxonomy;
- Identify the role (scope and scale) of existing laboratory systems in the model and diagrams; and
- Test the model and diagrams against each of the laboratory areas and other interested parties (IT, legal, QA, records management).

A high-level plan, showing the relationships, processes and data flows that describe a future state for the laboratory, should be developed. This should include an identified role for each of the laboratory systems and should clarify the specific functions of each. Any problems with laboratory terminology should be resolved. The plan should be tested by presentation and discussion with the interested parties.

Business plan development

- Quantify the benefits of the proposal, in particular productivity gains, ROI and knowledge management, and support these estimates with case studies;

- Undertake a risk assessment, paying attention to process, technology and people-related risks. Align the risk assessment to the set of user requirements; and
- Prepare, and include in the business case, a high-level implementation plan that addresses any specific requirements and/or risks that have been identified.

Quantitative benefits should be identified, along with all risks. An implementation plan should address known risks and/or potential problems, in particular the strategic approach to roll out, e.g. a progressive deployment, the composition of the project team, change management and user support.

Human factors

- What practical problems do laboratory workers experience with existing laboratory processes and data workflows?
- How well will laboratory workers accommodate change? and
- Are there any cultural, political or other internal relationships that could have an impact on the project?

Potential problems associated with change management show be identified. This may be at an individual level or at an organisational level.

Internal culture and technology adoption

The introduction of multi-user IT systems into organisations has a mixed track record. Multi-user systems are usually specified by a project team and often contain a number of compromises and assumptions about the way people work. High-level business objectives can therefore be put in jeopardy if users do not successfully adopt the new system. However, most case studies on electronic laboratory notebook implementations indicate a positive user take-up. This may be attributed to a growing understanding of aspects of technology adoption, originally reported by Everett Rogers in his book, *The Diffusion of Innovations,*[15] and developed further by Geoffrey Moore in *Crossing the Chasm*[16]. Moore's 'Chasm' (see Figure 7) is the gap between the early adopters and the mainstream market. The early adopters are a relatively easy market. Targeting them initially is important, but the next phase of the marketing strategy must target the conservative and pragmatic majority. The early adopters can play a central role in this. Since the electronic laboratory notebook (ELN) project team is likely to be formed from the early adopters, they can play a pivotal role not only in specifying and selecting a solution, but in articulating the rationale for the ELN, provide training and ongoing support to the conservative and pragmatic majority.

User adoption is often considered one of the most critical success factors of an IT project, and paying appropriate attention to user requirements will enhance the likelihood of success. Key to this is the recognition that people are more likely to comply with a request when:

- A reason is provided;
- There is give and take;
- They see others complying;
- The request comes from someone they respect or like; and
- The request comes from a legitimate source of authority.

Concerns about user adoption can be reduced by carefully choosing the project team to ensure that these criteria are addressed, rather than just announcing a new system and the training course schedule. Typically, putting a strong emphasis on user requirements and user adoption by engaging users throughout the process tends to brand the implementation as a 'laboratory' project, rather than an 'IT' project, and this can make it easier for scientists to accept the proposed change.

The Technology Acceptance Model[17] (see Figure 8) is an information systems theory that models how users come to accept and use a technology. The model suggests that, when users are presented with a new software package, a number of factors influence their decision about how and when they use it. The main ones are:

- Perceived usefulness (PU): 'The degree to which a person believes that using a particular system would enhance his or her job performance'; and
- Perceived ease-of-use (EOU): 'The degree to which a person believes that using a particular system would be free from effort.'

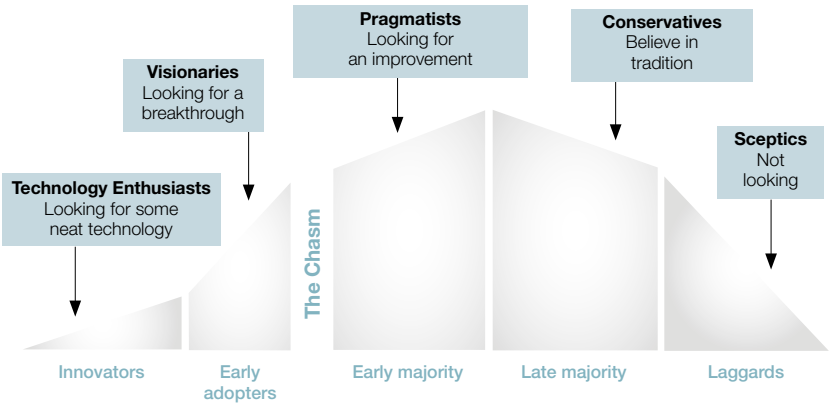
The technology acceptance model assumes that, when someone forms an intention to act, they will be free to act without limitation. In the real world there will be many constraints such as limited ability, time constraints, environmental or organisational limits, or unconscious habits that will limit the freedom to act.

Concentration on the positive aspects of 'usefulness', both to the organisation and to the individual, and 'ease of use' will help users develop a positive attitude. It is in this area that the early adopters can have a powerful influence on their conservative and pragmatic peers.

Technology considerations

Multi-user informatics systems are typically based on two- or three-tiered structures in which the application software and database may share a server or be located on separate servers, and the client-side software deployed on a local desktop, laptop or mobile device. Traditionally, the servers are based in-house, but hosted services (cloud/SaaS) are generating

FIG 7 Crossing the Chasm



Innovators	2 to 3 per cent	Technology enthusiasts: want to be first to try new technology; want one of everything.
Early adopters	10 per cent	Visionaries: able to align technology with strategic opportunities; willing to take risks; horizontally oriented.
Early majority	36 per cent	Pragmatists: cautious with risk and money; loyal; vertically oriented.
Late majority	36 per cent	Conservatives: opposed to discontinuous innovation; believe in tradition rather than progress.
Laggards	15 per cent	Sceptics: negative attitude towards technology; identify discrepancies between what's promised and what's delivered.

increasing interest, based on potential business benefits.

From the user perspective, the client-side options fall into two categories: thick client and thin client. The thick client is usually a substantial software installation on a local computer in which a good deal of the data processing is undertaken before passing the output to the database server. This has the advantage of distributing the total processing load over a number of clients, rather than the server, and may also allow a certain amount of personalisation of the client software to support individual users' needs.

The downside is that system upgrades can become time-consuming and potentially troublesome, depending on the local configuration – although centrally managed systems are now making thick client systems easier to deploy, maintain and support.

Thin clients typically access the application and database server(s) through a browser. No local processing power is used, so the server and network performance are critical factors in providing good performance. The use of a browser can significantly reduce deployment and upgrade costs, but may restrict or limit user configurability.

With regard to devices, successful deployments have been made with:

- Small form-factor PCs on the laboratory bench;
- 'Remote desktop';
- Citrix; and
- A KVM switch operating between a desk-bound processor unit with keyboards and screens on the desk and in the laboratory.

There is a growing level of interest in how consumer technologies can enhance the user experience of working with laboratory informatics tools. With their focus on sharing, collaboration, interaction and ready access to information, consumer technologies exhibit considerable synergy with the high-level criteria associated with current business requirements. Primarily, these focus on 'mobile' (portable devices), 'cloud' (access from anywhere), 'Big Data' (the need to be able to access and interpret vast collections of data) and 'social' (collaborative tools).

The big attraction of mobile devices for end users is portability. A common complaint in the transition from paper systems to electronic is the loss of portability of, for example, a paper lab notebook.

The form factor of a laptop computer goes part way to resolving this concern, but the emergence of compact, lightweight tablets holds far more potential. Although tablets are often considered to be 'data consumers' – great for reviewing data but less effective for data entry – careful design of the user interface can optimise their potential for narrow, dedicated functions.

Typically, mobile devices offer significant potential for accessing data from remote locations, or for capturing certain types of data in the field. The user experience can be enhanced by the use of mobile devices that feature simple, gesture-based interactions for on-screen navigation, consistent with typical consumer applications. Furthermore, the adoption of web technologies creates the opportunity to design a platform that supports all types of end-user devices, making critical laboratory data available anytime, anywhere, on any device in a global wireless and mobile environment.

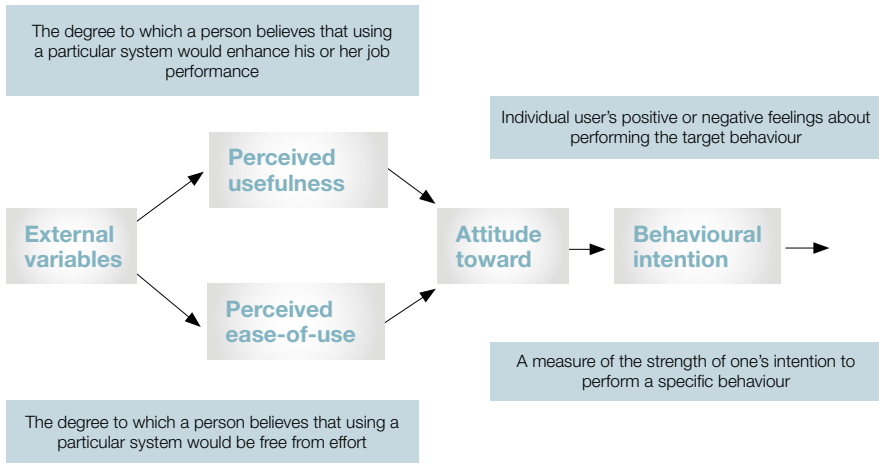
The adoption of mobile devices for informatics-based tasks raises a further question about how the host system is deployed and, in particular, how the mobile device communicates with the host. Synchronisation is one option, which has the advantage of not requiring remote connectivity, but it means that data must be held locally on the device. Wireless connectivity to a hosted system (SaaS or cloud) has the benefit of direct access to the system.

From a business perspective, the cloud offers an effective solution to the increasing demand for the implementation of collaboration tools across multiple departments, multiple sites and different geographies – including outsourced operations where the practicalities of deployment are largely limited to configuration rather than physical installation of hardware and software. The benefits of a thin client – access from anywhere, low start-up costs and centralised support – has both financial and functional attractions.

Pitted against this are concerns about access control, security, and data integrity.

Some informatics vendors already offer this type of service. Cloud services generally fall into one of two categories: public clouds and private clouds. Public clouds utilise a single code base for the service to multiple clients. The single code base limits customisation and integration, but helps keep costs down. A private cloud will typically offer a code base specific to an individual client, and will accommodate customisation and integration, but will normally come at a higher management cost.

FIG 8 Technology acceptance model



There is a growing level of interest in how consumer technologies can enhance the user experience of working with laboratory informatics tools

"

Chapter summary

The purchase and implementation of a laboratory informatics system represents a major cost to the laboratory. It also represents the start of a relatively long-term relationship with the vendor. Deploying a new system changes the working lives of laboratory workers and, as is the case with any significant change, planning takes on a critical role in the process. ■

KNOWLEDGE

Data analytics

This chapter takes the theme of knowledge management beyond document handling into the analysis and mining of data. Technology by itself is not enough – laboratory staff need to understand the output from the data analysis tools – and so data analytics must be considered holistically, starting with the design of the experiment

Data analytics is the term applied to the process of analysing and visualising data, with the goal of drawing conclusions and understanding from the data. Data analytics is becoming increasingly important as laboratories have to process and interpret the ever-increasing volumes of data that their systems generate.

In the laboratory, the primary purpose of data analytics is to verify or disprove existing scientific models to provide better understanding of the organisation's current and future products or processes.

Data mining is a related process that utilises software to uncover patterns, trends, and relationships within data sets. Although data analytics and data mining are often thought of in the same context, often in connection with 'Big Data', they have different objectives.

Data mining can broadly be defined as a 'secondary data analysis' process for knowledge discovery. It analyses data that may have originally been collected for other reasons. This differentiates it from data analytics, where the primary objective is based on either exploratory data analysis (EDA), in which new features in the data are discovered, or confirmatory data analysis (CDA), in which existing hypotheses are proven true or false.

In recent years, some of the major laboratory informatics vendors have started to offer data analysis and visualisation tools within their product portfolios. These tools typically provide a range of statistical procedures to facilitate data analysis; and visual output to help with interpretation. Alongside the integrated data analytics tools, more and more vendors offer generic tools to provide software that can extract and process data from simple systems through to multiple platforms and formats. The benefit of integrated data analysis tools is that they will provide a seamless means of accessing data, eliminating concerns about incompatible data formats. As with any other laboratory software, defining functional and user requirements are essential steps in making the right choice. Key areas to focus on are that the tools have appropriate access to laboratory, and other data sources; that they provide the required statistical tools; and that they offer presentation and visualisation capabilities that are consistent with broader company preferences and standards.

Data analytics plays an important role in the generation of scientific knowledge and, as with other aspects of 'knowledge management', it is important to understand the relationship between technology, processes, and people. In particular, staff need to have the appropriate skills to interpret, rationalise, and articulate the output presented by the data analysis tools. To take full advantage of data analytics, it should be considered as part of a holistic process that starts with the design of the experiment.

A quote attributed to Sir Ronald Fisher, ca 1938, captures this point: 'To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: He may be able to say what the experiment died of.' ■

Summary

In this guide we have attempted to coalesce much of the information required in order to design and implement as smart laboratory or, at the very least, to begin the process of laboratory automation. While it may seem like a challenging prospect, the underlying principles are simple and focused on crafting a strategy that will enable more productivity and insight to be generated from scientific research

The five nodes in Figure 9 represent a generalisation of the major knowledge processes, and it is quite evident that technology, in the form of the laboratory informatics tools, has an enabling role in a laboratory knowledge ecosystem.

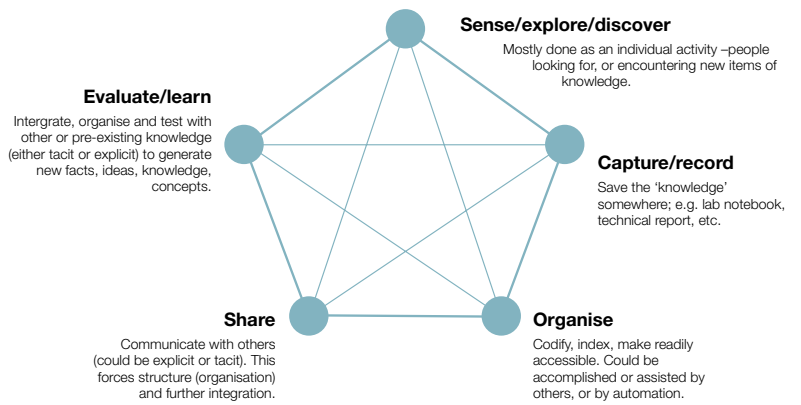
From laboratory informatics to knowledge management, technology is predicated on logical and systematic processes. But serendipity has always had a significant role in science. Many scientific advances have originated from 'what if' moments, chance observations, and things that went wrong. Failure often has more to teach than success!

With a growing emphasis on right-first-time, error-reduction, and productivity, it is a management challenge to take the time to review and assess successes and failures.

The role of the informatics tools within a smart laboratory, or 'knowledge ecosystem' (Figure 9) is important. They are strong in terms of capturing, recording, and organising data, and increasingly, they provide facilities for sharing information. But further opportunities arise with regard to evaluation and learning (data analytics) and experimental design. Further afield, they will contribute to predictive science.

Nevertheless there needs to be some space for 'right brain' thinking, alongside those systematic and structured approaches for increasing efficiency and productivity. Innovation depends on knowledge and understanding. Although technology can assemble and look after the data, making sense of it is down to human assessment

FIG 9 Knowledge processes



and understanding. In the main, so-called knowledge management 'solutions' are no more than data or information management solutions.

It's only when the human component is added that knowledge management can flourish, and even then, it needs the right environment – hence the concept of a 'laboratory ecosystem', or smart laboratory. Although management may want to see such an ecosystem, it can buy only the tools, it needs to create the right environment if the knowledge ecosystem is to be nurtured and cultivated.

The ecosystem is dependent on an open and collaborative culture and supportive leadership; not secrecy, discipline or rigid management. Participants need to opt in; not be forced in. One worry is that the digital revolution may be driving a lot of thinking to be 'digital', with the risk that random, analogue mindsets and gut feelings may be seen as irrelevant and inconsistent with modern concepts of science.

This way of looking at things may shed some light on why the early ELN market was sub-divided into different solutions for chemistry, biology and QA.

The risk for a multi-disciplinary laboratory that is looking to implement an ELN would be to adopt a one-size-fits-all approach. This could generate disaffection amongst users.

The current informatics market is moving towards more modular solutions, which have a generic core, and optional discipline-specific modules. This creates a better opportunity to find a single-source solution. The shared functions can be separated from the scientific functions which are closer to the heart and soul of the scientist's laboratory work. Shared functions would include such issues as document authoring, approval/witnessing, file and document management, and legal and regulatory compliance – all of which fall into the 'bureaucratic' category and which lend themselves to process improvement opportunities more readily than the scientific aspects. It may still be a one-size-fits-all approach, but it can be designed to accommodate the requirements of multi-disciplinary laboratories, and to standardise and improve common sub-processes, rather than making compromises.

Ideally, laboratory informatics tools should not be perceived as an intrusive bureaucratic process, but rather as something that facilitates the scientific method and doesn't intrude on the social and intellectual processes that are essential to the science. Achieving this objective is essential to joined-up science and to user acceptance, and is a responsibility that falls to management in its objective of building a smart laboratory. It requires a sympathetic view of the requirements of the different disciplines, and the way in which these functions are managed and provided for, even when organisational demands push for increased uniformity and consistency.

The concept of a smart laboratory will vary from organisation to organisation depending on the nature of its business, and the technological choices it makes. Discovery and development are increasingly recognised as two steps in a holistic product life-cycle process rather than stand-alone functions.

The focus of this guide has been on technology, with due consideration to the laboratory processes to which it can be applied. It has also touched on some aspects of culture and technology adoption, but it must be remembered that user acceptance is a critical success factor in almost every system or project.

Technology on its own cannot overcome challenges in the laboratory. The take-home message is that to become 'smart' the lab users' and managers needs to understand its role in the organisation's end-to-end business processes and optimise its technologies to fulfil those requirements. ■

References

1. The Gartner Hype Cycles: www.gartner.com
2. CENSA: The Collaborative Electronic Notebook Systems Association
3. Good Manufacturing Practice (GMP): The US has one set in the Federal Register 21 CFR, and the EU has its own, as do other geographical areas and organisations like the Organisation for Economic Co-operation and Development (OECD)
   US FDA, 21 CFR Part 211 Current Good Manufacturing Practice for Finished Pharmaceuticals, 2005, FDA: www.fda.gov
   European Union Volume 4: Good Manufacturing Practices – Medicinal products for human and veterinary use, 1998, 153 pages, incl. Annex 11 covering computerised systems
4. Good Laboratory Practice (GLP): The US has one set in the Federal Register 21 CFR; EU has its own, and also other geographical areas
   US FDA, 21 CFR Part 58 Good Laboratory Practice for Non-Clinical Laboratory Studies, 2005, FDA: www.fda.gov
   European Union, Council Directive of 7 June 1988 on the inspection and verification of Good Laboratory Practice (GLP) (88/320/EEC)
   European Union, Council Directive – of 24 November 1986 - 86/609/EEC – on the approximation of laws, regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes
5. US FDA, 21 CFR Part 11 Electronic records; electronic signatures, 1997, FDA: www.fda.gov
   OECD Series on principles of good laboratory practice and compliance monitoring number 10
6. European Union Volume 4: Good Manufacturing Practices – Medicinal products for human and veterinary use, 1998, 153 pages, incl. Annex 11 covering computerised systems
7. PIC/S, PI 011-03 Good practices for computerised systems in regulated 'GxP' environments. 25 September 2007, PIC/S: www.picsscheme.org
8. GAMP 5 (Good Automated Manufacturing Practice) Guide: A Risk-Based Approach to Compliant GxP Computerized Systems, February 2008, International Society for Pharmaceutical Engineering (ISPE), Fifth Edition, ISBN 1-931879-61-3: www.ispe.org
9. Good Automated Manufacturing Practice Guidelines version 5, International Society for Pharmaceutical Engineering, Tampa FL, 2008
   McDowall, R.D., (2009) Spectroscopy Focus on Quality, p23
10. Using Electronic Records in Patent Proceedings, article by Damien McCotter and Peter Wilcox. Originally published in Managing Intellectual Property's World IP Contacts Handbook, 14th edition, 2007. Available at www.mondaq.com
11. IP Expert Advice: Tips on creating a lab notebook that contains 'convincing evidence': www.edn.com/article/CA6445886.html?industryid=47048
12. Admissibility of Electronic Records in Interferences, Bruce H. Stoner Jr., Chief Administrative Patent Judge, www.uspto.gov/web/offices/com/sol/og/con/files/cons119.htm
13. Private communication: Colin Sandercock (Perkins Coie LLP) September 2011
14. The ABCs of Electronic Signatures, Nettleton, D., Lab Manager Magazine, 9 September 2010: www.labmanager.com/?articles.view/articleNo/3800/title/The-ABCs-of-Electronic-Signatures
15. Rogers, E. M., Diffusion of Innovations, The Free Press. New York
16. Moore, G. A., Crossing The Chasm, Capstone Publishing
17. Bagozzi, R. P., Davis, F. D., and Warshaw, P. R., (1992). Development and test of a theory of technological learning and usage. Human Relations, 45(7), 660-686
18. Multi-Ontology Sense Making, David Snowden, http://cognitive-edge.com/uploads/articles/40_Multi-ontology_sense_makingv2_May05.pdf

Further reading and websites

Stafford, J. E. H., (1995) Advanced LIMS Technology: Case studies and business opportunities, Springer

Christensen, C. M., (1997) The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail, Harvard Business School

Segalstad, S. H., (2008) International IT Regulations and Compliance: Quality Standards in the Pharmaceutical and Regulated Industries, Wiley-Blackwell

McDowall, R. D., (1987) Laboratory Information Management Systems, Sigma Press

Laboratory Notebook Guidelines: BookFactory, LLC, 2302 S. Edwin C. Moses Blvd, Dayton, OH 45408. Available at www.bookfactory.com

Mahaffey, R. R., (1990) LIMS: Applied Information Technology for the Laboratory, Nakagaw

Sellen, A. J., and Harper, R. H. R., (2003) The Myth of the Paperless Office, The MIT Press

Franklin, C., (2003) Why Innovation Fails, Spiro Press

Kanare, H. M., (1985) Writing the Laboratory Notebook, An American Chemical Society Publication

eOrganizedWorld: www.eorganizedworld.com

Free online LIMS training courses: www.LIMSuniversity.com

The Generally Accepted Recordkeeping Principles: www.arma.org/garp/index.cfm

Independent, non-commercial LIMS user's group: www.LIMSforum.com

Industrial Lab Automation: www.industriallabautomation.com

The Institute for Laboratory Automation: www.institutelabauto.org

The Integrated Lab: www.theintegratedlab.com

Journal of Information & Knowledge Management (JIKM): www.worldscientific.com/worldscinet/jikm

LIMSfinder: www.LIMSfinder.com

NL42 Consulting: www.NL42.com

Online encyclopaedia for laboratory, scientific and health informatics: www.LIMSwiki.org

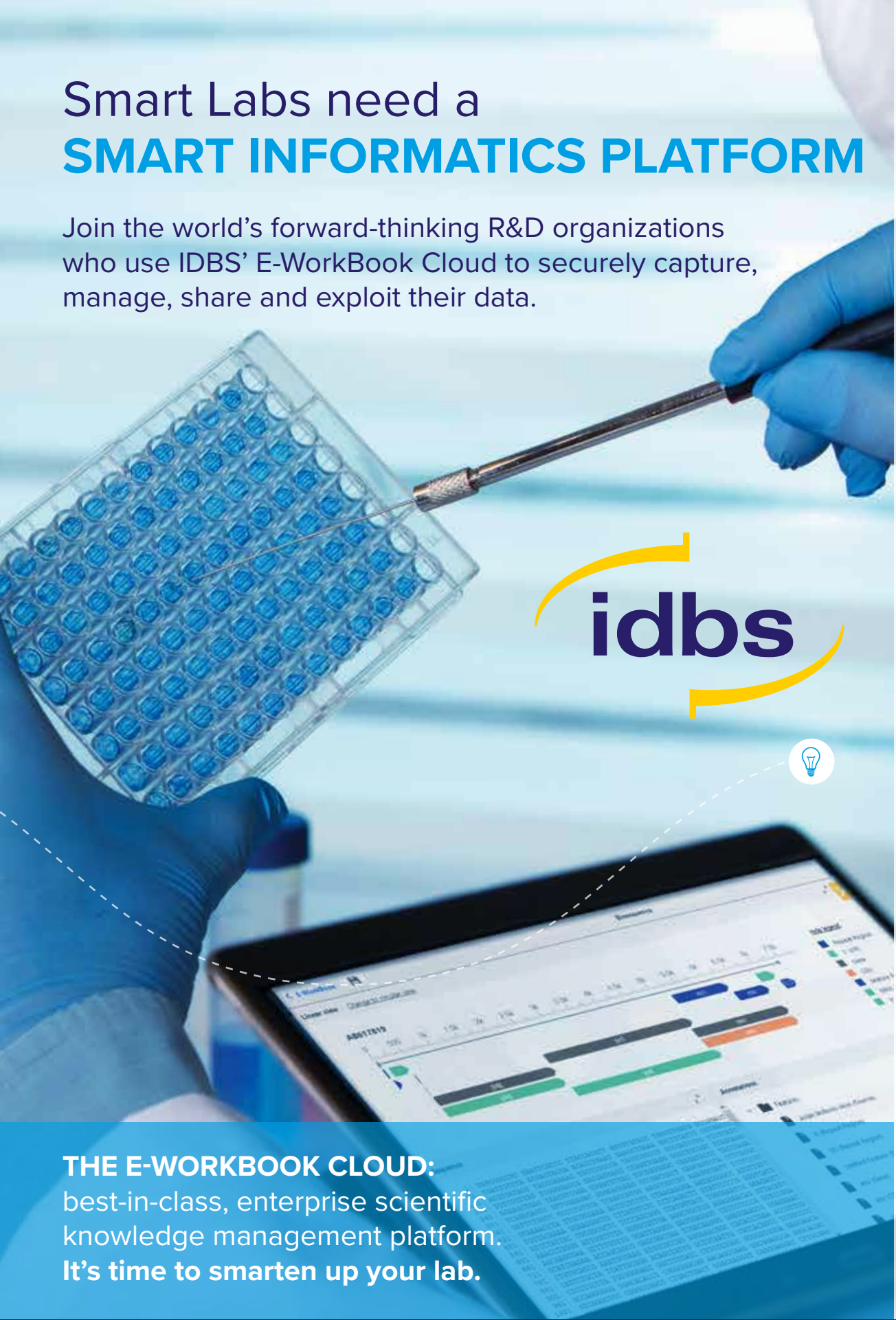
PDF/A standard: http://en.wikipedia.org/wiki/PDF/A

Scientific Computing World: www.scientific-computing.com

Segalstad Consulting: www.limsconsultant.com

Smart Labs need a **SMART INFORMATICS PLATFORM**

Join the world's forward-thinking R&D organizations who use IDBS' E-WorkBook Cloud to securely capture, manage, share and exploit their data.



PLATFORM MODULES

- ELN
- ADVANCE
- CHEMISTRY
- BIOLOGY
- INVENTORY
- CONNECT
- REQUEST
- INTEGRATIONS

THE E-WORKBOOK CLOUD:
best-in-class, enterprise scientific knowledge management platform.
It's time to smarten up your lab.

WWW.IDBS.COM

idbs

thermo scientific

Scale up your expectations, Chromeleon CDS does more.

From workstation to global deployment, Thermo Scientific™ Chromeleon™ CDS scales to any business – from research through development to QC – delivering proven productivity gains and easy tracking and defense of your data. One cloud-ready CDS that does it all – all backed by a product team dedicated to keeping you one step ahead of the next regulatory or market change.

Learn more at thermofisher.com/Chromeleon

ThermoFisher SCIENTIFIC