Chem Soc Rev

RSCPublishing

REVIEW ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Soc. Rev.*, 2013, **42**, 8157

Received 30th March 2013

DOI: 10.1039/c3cs60122f

www.rsc.org/csr

Laboratory notebooks in the digital era: the role of ELNs in record keeping for chemistry and other sciences

Colin L. Bird, Cerys Willoughby and Jeremy G. Frey*

Egyptian evidence of scientific records dates back almost 50 centuries. In more recent times da Vinci and Faraday provide role models for scrupulous recording of ideas, observations, and conclusions. Their medium was paper, but despite the quality of their notebooks, we cannot turn the clock back. Our primary purpose is to review the influences of the digital era on scientific record keeping. We examine the foundations of the emerging opportunities for preserving and curating electronic records focussing on electronic laboratory notebooks (ELNs), with an emphasis on their characteristics and usability.

Introduction

The origin of science itself is a subject for philosophical discussion; so dating the first recording of a scientific endeavour must almost inevitably also be a matter for speculation. However, we do know of an instance of recorded occupational medicine that has been attributed to the Egyptian architect, physician, and statesman Imhotep (27th century BCE), based on a study of the Edwin Smith papyrus.[1]

The History of Science article in Wikipedia covers the development of science and scientific methodology from the ancient civilisations through to the modern era, and includes several examples of the recording of information in a manner that we could regard as scientific. Observations were represented numerically as well as in narrative form, for example the recording of astronomical information in Mesopotamia.[2] There is also historical evidence of the reuse and repurposing of

Chemistry, Faculty of Natural and Environmental Sciences, University of Southampton, Highfield, Southampton, SO17 1BJ, UK. E-mail: j.g.frey@soton.ac.uk

Colin L. Bird

Having obtained his BSc and PhD in Chemistry at the University of Southampton, Colin Bird joined IBM UK Laboratories. After contributing to IBM's electrochromic display technology, he transferred to the IBM UK Scientific Centre to develop advanced image and visualisation applications. His work on content-based image retrieval led to a one-year secondment in 1999 back to the University of Southampton. On returning to IBM, he was involved in various aspects of information management, specialising in classification and metadata, and became an information architect. When he left IBM, he resumed his collaboration with Professor Jeremy Frey on e-Research projects, which began in 2000 as an industrial partner for the CombeChem project.

Cerys Willoughby

Cerys Willoughby obtained her BSc in Geology from the University of Wales, Aberystwyth and her MSc in Environmental Sciences from University of Wales, Swansea. After spending some time teaching teenagers and adults, she joined IBM UK Laboratories as a software engineer. Working in a variety of roles in the company she specialised in usability and information architecture. She began collaborating with Professor Jeremy Frey in 2007 as a guest lecturer on usability and accessibility for an e-Research course and is currently undertaking a part time PhD at the University of Southampton.

ancient records, for example the emergence in 15th century Italy of astronomical understanding based on the creations of Greek science.[3]

A comprehensive study of why humans preserve records of their various activities would be the province of anthropology; in this paper we focus on why and how record keeping is an integral part of scientific methodologies. The reasons we might advance for capturing a record of scientific and technical activities range from: aiding our thought processes; through enabling communication and collaboration; to the protection of intellectual property and even compliance with official requirements.

Shankar adopts a more philosophical approach, in which he decomposes record keeping into three constituent acts: information synthesis, formalization, and tool-making. Using this analysis, he develops a model in which personal and collaborative information management is a product of recording and the progressive addition of further information, narrative, and annotation.[4]

This model also, in some respects, exemplifies the evolution of scientific record keeping. The paragons of recording – Leonardo da Vinci, Michael Faraday, and others – kept diary-style notebooks. Paper was the only medium available, so was also the means for communicating with fellow scientists, either by personal letter or by publication in a learned journal. Data, obtained by observation and experiment, was also preserved in the notebook, as was information derived by analysis of that data. The advent of computers and digital storage has led to significant changes in both capability and approach: communications have become more open and less formal; collaboration has become more extensive.

However, in more recent times, scientific research has acquired a competitive element: researchers value ownership of their data as their intellectual property, and can be reluctant to share information until papers have been published and/or data is no longer commercially sensitive. This protective attitude sometimes extends into unwillingness to record thoughts and ideas, for fear of their misuse by others: easy communications are not always beneficial for scientific and technical record keeping. There is a view that "if I say too much, others will find out more than I want them to and/or sooner than I want them to"!

Jeremy G. Frey

Jeremy Frey obtained his DPhil on experimental and theoretical aspects of van der Waals complexes, in the PCL, Oxford, followed by a NATO/SERC fellowship at the Lawrence Berkeley Laboratory. In 1984 he took up a lectureship at the University of Southampton, where he is now Professor of Physical Chemistry. His experimental research probes molecular organization in environments from single molecules to liquid interfaces using laser spectroscopy from the IR to soft X-rays. He investigates how e-Science infrastructure can support scientific research with an emphasis on the way appropriate use of laboratory infrastructure can support the intelligent access to scientific data.

Notwithstanding such reservations, we cannot turn the clock back. Computing and computers are now pervasive in science and technology. Computers now control most instruments; computational methods, such as simulations, provide *in silico* tools and techniques; and computers enable us to capture, analyse, and annotate our data. As experimentalists we often do not even see the raw data, as significant pre-processing of the data takes places automatically within the "black box" that is the "equipment". Each year computing facilities become more powerful, almost a necessity just to keep pace with the expanding volume of data. In the 21st century we cannot even hope to keep track of our results and other notes without digital support. Moreover, when data is semantically annotated, programs can reason and actually assist with the implementation, and undertaking of the science itself.

As researchers come to terms with the implications of such advances, particularly cheap and extensive digital storage, the scientific community encourages researchers to preserve all their records. Public funding of research now commonly comes with a mandate to store data in open repositories. Significantly for this paper, we are hearing more often the old mantra: *show your working*!

In this paper, we examine the foundations of record keeping, as the basis for considering the changes that have come since we entered the digital era. These transformations affect both the range and the nature of the records we keep and, in some cases, the reasons for preserving those records. We explore two particular manifestations of adapting to the digital era: the sufficient and appropriate curation of digital records, and the expanding use of electronic laboratory notebooks (ELNs). We present the results of a recent survey of the literature relating to ELNs.

From this survey and our own work in the area of the Chemical Semantic Web and ELNs it is clear that if we are to meet the challenges of record keeping in the digital era, especially the fears of some practitioners, the proper consideration of usability becomes an imperative. We appraise some of the issues arising from usability studies, in the light of which we consider in the final sections a selection of pointers to the future, such as open access, data publication, and the influence of the Semantic Web.

Scientific recording: role models

The perennial fascination with Leonardo da Vinci makes the archetypal polymath an ideal role model to begin with. His notebooks contain, among other items, scientific diagrams and, significantly, his thoughts, thus comprising a notable endowment to subsequent generations. The Wikipedia article about Leonardo expresses the view that the structure of his notes suggests that he intended those notes for publication.[5]

Michael Faraday (1791–1867) is still renowned for his scrupulous recording of all aspects of his research:

Faraday's laboratory notebooks are also remarkable in the amount of detail that they give about the design and setting up of experiments, interspersed with comments about their outcome and thoughts of a more philosophical kind. All are couched in plain language, with many vivid phrases of delightful spontaneity. . .[6]

Charles Darwin (1809–1882) used his notebooks for his observations in the field and for theoretical speculations: the manuscripts are available online.[7] The extensive collection of papers left by Albert Einstein (1879–1955) is also available online, providing insights into his thoughts and calculations.[8]

One of the characteristics common to all four role models is their recording of thoughts as well as observations and calculations. They show their working and thinking and thus enable other scientists to understand the provenance of their conclusions and to reuse their findings. As many have noted, the progress of science depends on individual scientists building on the results already produced by others; anything that makes this reuse easier and more reliable is clearly to be encouraged.

The nature of the scientific record

Thirty years ago, Eisenberg cited several illustrations of the importance of keeping good records, going on to proffer general guidelines for maintaining a laboratory notebook.[9] The summary she gives is as valid today as it was when she wrote the following words:

The uses of laboratory notebooks are not limited to legal issues. They are vehicles for organizing and focusing the thinking of the writer, as well as being receptacles for detailed procedural information that might not be available in highly compressed journal articles. Finally, they may serve not only the researcher or inventor but also the public. If properly maintained, they are a record of success and failure, a safe-guard against error and carelessness in such important areas as the testing of drugs and chemicals.

The online Oxford Dictionary defines the *scientific method* as follows:

A method of procedure that has characterized natural science since the 17th century, consisting in systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses.[10]

In slight contrast, Merriam-Webster online gives the following definition:[11]

''principles and procedures for the systematic pursuit of knowledge involving the recognition and formulation of a problem, the collection of data through observation and experiment, and the formulation and testing of hypotheses''

The latter definition is perhaps closer to ''scientific thinking'' and less ''methodological'', but what these both have in common is that evidence and the recording of evidence are fundamental to the scientific thinking and method and are the basis of reproducibility: *The integrity of science as a discipline rests on the ability of scientists to reproduce the claims of others.*[12]

Fundamental to all practitioners of scientific research is that results should always be open to testing and be capable of replication. Before publishing an article, the journal *Organic Syntheses* requires each procedure to be reproduced, as described, in an independent laboratory. We can expect the record of any given endeavour to comprise data and information in most or all of the categories: ideas, conjectures, plans, details for setting up equipment, methods, observations, results, analyses. In the digital era, we might also look for processes and workflows, as exemplified by the myExperiment repository.[13]

Within the data-information-knowledge-wisdom hierarchy (Ackoff, 1989),[14] *a research chain of knowledge is located between the information and knowledge levels. It is more than an information bit because it includes the interrelation between several information bits and their appropriate documentation.*[15]

Frey and Bird discuss the data-information-knowledge-wisdom (DIKW) pyramid in the chemistry context in a review of Chemical Information.[16] Regrettably, failure to record the science properly can sometimes give rise to public concern, and embarrassment for those involved. Ince's article about the so-called 'Duke University scandal' strongly demonstrates the importance of provenance information for both audit and reproducibility.[17] The article attributes the affair in part to *sloppiness in data curation and software storage*.

The ClimateGate affair, as it came to be known, arose from the publication of e-mails and other documents that researchers at the Climatic Research Unit, University of East Anglia, had thought were private. The BBC News item was one of many media articles exploring the consequential demands for greater public access to research data, for scientists to ''show their working'' to the public as well as when subject to peer review.[18]

Preserving the record

Paper continues to be attractive for capturing and preserving the records of human activities, scientific and otherwise. Paper records are portable, do not require a power supply, and can be stored securely. The disadvantages are the risks of loss or destruction, and the complications with retrieving material, particularly data, for reuse.

For scientists and practitioners in other spheres of technical activity, publication is the preferred mechanism for enabling wider access to their material and providing the appropriate recognition for their work, although publication does not constitute a full archival record. However, some of this information, such as reports, conference proceedings, theses, and translations, remains difficult to identify and access. Library and information scientists refer to such material as *grey literature* and have formed organisations such as GreyNet to deal with the distribution of, and access to, such material.[19]

Historical instances of difficulties with the granting of patent rights were the basis for Eisenberg's advice about the importance of good records.[9] Concerns about intellectual property (IP) rights are as alive today as they were then, especially in those commercial arenas where research is highly competitive: drug discovery is a prominent example.

Concerns about protecting IP are not confined to textual material but extend to raw and derived data as well. The competitive aspect of research leads to reluctance to share data until papers have been published and/or data is no longer commercially sensitive.[20] Downing *et al*. conducted a survey of research chemists at Cambridge and Imperial College, intended primarily to investigate the factors that would encourage respondents to share research data in an open access repository. The responses showed a clear reluctance to allow immediate open access to research results, permitting only other group members to see information before publication. Significantly, they also found a tendency to store data as hard copy.[21]

Related reasons for preserving research records arise from requirements to comply with regulatory conditions. Pharmaceutical companies testing new drugs are required to maintain comprehensive audit trails to guarantee the provenance of their test results. For academia, the research councils now mandate the preservation in public repositories of any data that is the product of publicly funded research: such data is a public asset and should be available for verification and reuse. In support of this initiative, JISC (formerly the Joint Information Systems Committee) is currently funding a series of research data management infrastructure projects.[22] It is our contention that good practice in record keeping in the laboratory, combined with the most useful aspects of the digital world, will make the managing of research outputs much easier. Once research outputs are in a digital repository, standards (such as CERIF, OAI-PMH, ORE) are now in place to replicate, discover, and cite these resources; less advanced are the standards around the data on which these outputs depend. JISC, the UK Research Councils, and the Coalition for Networked Information (CNI) are all working towards reducing the barriers to improved scientific reporting.[23]

The potential for collaboration with other researchers has always provided an incentive for careful preservation, albeit subject to caution regarding the extent of the information shared. Latterly, the increase in the number of larger-scale and multidisciplinary projects has led to more preservation of data and other information in open and shared repositories.

Having considered the motivations, we now explore the mechanisms for preserving records of research. Discussions have largely focussed on ''Publication'' (in formal academic journals as opposed to the more general media or the monograph/book culture of the humanities) as this continues to be a prominent medium for sustaining the outcomes of scientific and technical endeavours.

Ball appraises the reasons for the success of journals as a means of scholarly communication, citing, amongst other reasons, raising awareness, protection from plagiarism, and permanent access.[24] However, he then argues that scientific journals are no longer capable of supplying all the information needed to ensure reproducibility: it is necessary also to provide the underlying data. Recognition of the need to preserve – and to make available – the full research record is one of the key consequences of entering the digital era. Bachrach supports this thinking, calling for electronic dissemination that includes the data.[25] His article is motivated by the difficulties of accessing the growing body of chemistry literature and by mitigating the rising cost of journals. In 2011, Lang and Botstein took the unusual step of publishing as supporting information a scanned copy in PDF format of the complete laboratory notebook, thus describing in full the work comprising the study reported in the paper.[26] MacNeil used this example of open science to draw a comparison between the paper and electronic views of laboratory notebooks, unsurprisingly availing himself of the opportunity to extol the virtues of the electronic form, and the iPad ELN in particular.[27]

How then do we preserve the supporting data in forms that simplify and encourage its reuse and repurposing? Traditionally, laboratory researchers have chosen flat files for data storage: they are simple to use; they can be copied to portable media for transferring elsewhere; in many cases laboratory instruments record their output in flat file format. The disadvantages of the flat file format are significant: they are difficult to control; items can be difficult to find and recover; assuring the integrity and provenance of the data is, to say the least, problematic. A relational database management system (RDBMS) or a laboratory information management system (LIMS) is at least capable of overcoming all of these disadvantages. For that reason, commercial organisations, at least for their analytical laboratories, in particular favour the option of management systems: audit trails exist and the requirements of due diligence can be met. Frey argues the need for solutions that cover the middle ground between uncontrolled flat files and the relatively inflexible management systems, suggesting that the Semantic Web provides the technology required in the laboratory environment.[28]

With the digital era have come various embodiments of the electronic laboratory notebook (ELN) as vehicles for preserving all parts of the research record in a consolidated form. We consider ELNs in greater depth in a later section: the electronic laboratory notebook (ELN): a literature survey.

Digital technology has transformed how we as individuals handle and manage information, particularly with the advent of *smart* devices. A team of archaeologists at the University of Southampton have experimented with using an iPad™ rather than a notebook when conducting archaeological excavations. The iPad enabled the archaeologists to capture information, such as audio and video, whilst walking around the site. They could then share that information in the form of a digital site tour. However, the archaeologists do not view the iPad as a replacement for notes, because adding text is difficult and they were unable to annotate material as they would in a paper notebook.[29]

Ensuring the availability of provenance information brings about a need to safeguard the quality and accessibility of the information and data, a process that we consider in the following section: Curating the record.

Curating the record

According to the Digital Curation Centre (DCC): *Digital curation involves maintaining, preserving and adding value to digital research* *data throughout its lifecycle*. This definition is followed by a list of the steps that comprise the digital curation lifecycle, a comprehensive view that serves to emphasise the importance of curation in scientific and technical record keeping.[30] Perhaps the most basic reason for curating records is to minimise the loss of information over time, a process illustrated very well, if somewhat wryly, by Fig. 1 in the paper by Michener *et al.* about ecological metadata.[31] The same paper also includes a telling justification:

> *The most important reason to invest time and energy in developing metadata is that human memory is short. If data are to undergo any secondary usage, then adequate metadata will be required even if that secondary usage consists of reuse by the data originator*.

The concept of metadata is not new; neither is metadata the only product of curation. The ingredients of curation are data and information (*digital objects* in the DCC view), together with descriptive and other metadata needed for effective and efficient access. However, the ingredients require careful preparation and presentation: in reality, curation will be concerned with a *package*, of which the data itself is but one component. The Open Archival Information System (OAIS) model, as described by Ball, is one example of an information package.[24]

Despite the importance of metadata, few authors are willing to attempt a rigorous definition of the term. The commonly accepted definition, that metadata is "*data about data*", soon runs into difficulty. Pancerella *et al.* assert: "*such a definition is very dependent on one's perspective*."[32] A full discussion of the nature of metadata is outside the scope of this paper, which is concerned with the capture of metadata as an aspect of record keeping. Although the majority of researchers would readily acknowledge the value of curation, many can feel discouraged by the effort required to curate their data properly.[20] However, designing curation into experiments and capturing metadata *at source* can mitigate the burden of curation.[33]

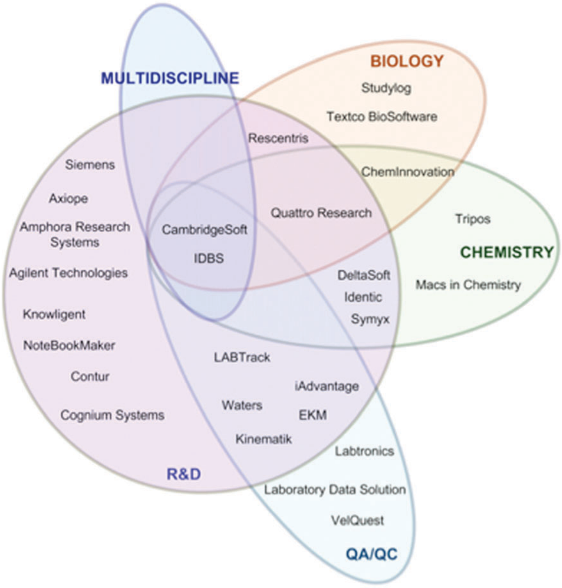

Fig. 1 Electronic laboratory notebook primary market audience, reproduced with permission from Anil Rattan.[58]

The electronic laboratory notebook (ELN): a literature survey

Three themes characterise the general discourse regarding ELNs: whether they represent evolution or revolution; the replacement of paper notebooks; and, albeit to a lesser extent, the pros and cons. Interestingly, the technology for implementing ELNs appears not to be a significant issue.

Williams *et al.* provide comprehensive guidance regarding the expected content, organisation and format of a paper notebook, together with extensive advice about recording experiments, from planning through running to data analysis and conclusions. Their treatise also includes a brief introduction to ELNs and ends with a discussion of intellectual property issues.[34]

In 1994, Borman took the view that ELNs could revolutionise how scientists record their research, manage their data, and share their information with others.[35] Recently, Lass adopted a more cautious view when discussing best practices for implementing ELNs:[36]

> *If not done correctly, moving from paper to an ELN will be perceived by scientists as a revolutionary activity. When the outlined process is followed, daily routine will be fully mapped to the ELN functionality, enabling scientists to continue documenting their experiments with minimal interruption. The movement to the ELN will be evolutionary and not revolutionary*.

According to Hice, there is no single definition of an electronic notebook, owing to differing requirements in different areas: he gives specific consideration to instrument interfacing. His view is that ELNs will evolve to meet market demands and that the *current line of demarcation between the different flavors of electronic laboratory notebooks may be a moot point one day*[37] with convergence with other forms of record keeping in a digital space being the highly likely outcome of current research and commercial efforts (for example some organizations can use Microsoft SharePoint for their record keeping requirements and more open collaborations often make use of Google docs).

Early opinions on the replacement of paper notebooks were comparatively radical: a 1998 study by the Collaborative Electronic Notebook Systems Association (CENSA) gave a list of reasons why *paper notebooks are obsolete.*[38] A 2003 editorial in Drug Discovery Today gave reasons for moving from paper to ELN;[39] other articles published at around the same time continued the radical view, one describing paper as *fundamentally flawed in the ability to share and manage data.*[40] Mullin concluded: *Once researchers are forced to use ELNs, they will likely never go back to paper – even if they are allowed.*[41]

Taylor acknowledges that CENSA was ahead of its time, but credits it with providing the first definition of an ELN: it is notable that the CENSA definition would still be acceptable today. He also includes a timeline showing the evolution of ELNs from before 1990 through 2010, going on to outline the benefits of ELNs from the perspective of a scientist's desktop.[42]

In 2007, Du and Kofman published a technology review, using a measured comparison, based on return on investment (ROI) calculations, of paper and electronic notebooks to produce a list of requirements.[43] A year later, Bruce made the interesting recommendation *that adoption of an ELN be voluntary — certainly in the pilot phase, but often in roll-out, too. This really forces an ELN to prove value to those that will use it, and demonstrates faith in people to decide what works best for them.*[44] In October 2011, MacNeil blogged a comparison, citing a PLoS One paper published with an entire paper notebook as supplementary data, and making a sound case for the ELN in collaborative research. His reasons included flexible organisation, linking to other aspects of the experiment, and sharing with both the team and the wider scientific community. Unsurprisingly, he also used this opportunity to extol the virtues of the iPad ELN.[27] In a similar vein, Elliott demonstrates how paper notebooks inhibit knowledge sharing. His Webinar sets the scene by outlining the characteristics of paper notebooks, and then proffering a view of knowledge management that involves ''a cultural migration to sharing, reusing and creating knowledge''. His examination of explicit and tacit knowledge, and the four possible transitions between these types, leads him to conclude that paper notebooks obstruct all four such transitions, whereas ELNs assist them.[45]

In the same year, Butler considered the pros and cons of electronic notebooks.[46] Successful adoption will undoubtedly depend on meeting the personal as well as the technical needs of researchers, as illustrated by an ethnographic study of scientific record keeping.[4] Among its conclusions were the words: *The experiences of the group of scientists studied in this article suggest that standardization of data entry is not the only aim of scientists when creating records*. The empirical evidence supports the stronger view that standards are far from being the top priority of those creating records in the laboratory.

Although CENSA is now inactive (according to Consortium-Info.org), the study cited earlier reported research into the functional and business requirements for R&D teams to be effective when using systems such as electronic notebooks. The article includes a cost-benefit analysis with some considerations that are still pertinent, although its observations about the web would not withstand scrutiny today.[38] It also takes a more relaxed view of IP issues than would come to be the case in later years.[47] We return to such considerations in the Regulatory and legal issues section, while noting here Elliott's conclusions from his review of the state of the ELN market in 2007:

> *The past year has seen most ELN installations migrating away from the 'hybrid' model that used printed and hand-signed pages for intellectual property protection. A fully electronic installation is now standard in the vast majority of large biopharmaceutical companies. PDF renditions are electronically signed and stored in a content management archive for long-term retention*.[48]

Such representations of ELN content do not create a versatile data repository, despite their importance for legal compliance, as it is difficult to extract data for reuse from the PDF format. Although it is possible to obtain some information with text-mining techniques, PDF renditions are an example of publication causing a loss of usability. The original data file is a much more useful resource.

Several writers have reviewed the evolving use of ELNs by pharmaceutical companies.[49–51] Recently, Kopach and Reiff have considered how ELNs can facilitate the calculation and reporting of green chemistry metrics associated with the development of new drug candidates.[52] This is an example of how new services can be built on top of digital data in a way that would be far too time consuming to do based on traditional paper records.

Characteristics of ELNs

Taylor has published two articles that evaluate the characteristics of ELNs. The first appraised their use in chemistry and biology, in which he predicted that eventually, all R&D scientists would use ELNs to preserve the records of their research.[53] Taylor's more recent assessment includes a claim that companies report large productivity gains as a result of the ability to reuse or repurpose details of experiments recorded previously in their ELNs.[42] However, hard economic data is difficult to come by.

In a somewhat different approach, Labtronics Inc. have published an eBook entitled *ELN and the Paperless Lab.*[54] This form of review comprises a set of articles, in their original form, that collectively consider the role, objectives, choices, potential benefits, and other considerations, such as configuration, connectivity, and security. Each essay can be read as a separate article, in which its conclusions are specific to its focus area.

The Du and Kofman technology review considers the processes that ELNs require for regulatory compliance in the pharmaceutical industry. They provide flowcharts for recording and reviewing an experiment and for good practice approval. They also present an interesting diagram showing workflow and data exchange points between the processes associated with an integrated laboratory and with the ELN.[43] This review accentuates the importance of proper control and management, a point endorsed by Elliott in an article collating input from several companies, showing that both knowledge management and productivity can benefit from using an ELN. He notes that careful and methodical project management leads to the *greatest positive impact of the technology and the highest levels of user satisfaction.*[55]

A Bristol-Myers Squibb case study proposed a five-step screening process for selecting a vendor ELN. This process includes a vendor questionnaire although, unsurprisingly, the key step is the definition of requirements: Fig. 1 in the article depicts the full composition of the questionnaire.[56] Previously, some articles had raised questions about the use of electronic notebooks for QA/QC, owing for example to legal and process issues and to the more formal structures required for QA/QC, but recently Metrick was able to take the view that ELNs could be suitable for such activities.[57]

Rubacha *et al.* have also published a review of 35 ELNs that are in the market today.[58] Of particular interest is their diagram showing the primary market audience of each ELN, which we reproduce with their permission as Fig. 1. This diagram uses overlapping ellipses for R&D, biology, chemistry, QA/QC, and – significantly – multidiscipline. Although the latter ellipse contains only two vendor ELNs, they begin their conclusions with the assertion: *Within the past 2 years, the overall landscape of available ELN solutions has evolved toward a new ideology; that of multidiscipline*. They base that view on the number of ELN solutions in their survey that they consider have general or some cross-functionality. Their categories were R&D, biology, chemistry, and QA/QC.

Elliott, in his annual review of the ELN market, offers an individual perspective on the development of commercial ELNs from the late 1990s. He views the evolution in terms of four phases, which he characterises as the following types:

- Basic data capture
- Specific solutions
- Expanded capabilities
- Converging functionalities

Looking to the future, he foresees a more holistic approach, with an information architecture that meets *increasing demands to bridge information silos across departments.*[59] This view follows naturally from his earlier presentation about the contribution of ELNs to knowledge sharing.[45]

In a way the antithesis to this sharing of information, that is knowledge protection, arose from intellectual property concerns and led to the emergence of hybrid digital–paper systems, which rely on printing final data to comply with patent requirements. For example, Schering AG developed a hybrid ELN system that used proprietary tools for authoring and document management, which they presented to the 2004 meeting of the American Chemical Society Division of Chemical Information (CINF in 2004). In the abstract of their talk they said:

The legal, archived version of completed experiments is printed to paper where it is signed and witnessed. There is not as yet significant case history in the US to support e-records used in patent litigation.[60]

At around the same time, Kihlén published two editorial articles about the Biovitrum ELN, which also produces printed records that are subsequently signed and witnessed for IP protection.[39,61] The 2003 paper makes the point that electronic notebooks do not obviate the need for paper, giving both legal and archival reasons. The same paper lists the requirements for a hybrid ELN, noting that successful solutions are as much about the organisation of work and processes as the functionality of the software.

Taylor's recent book chapter about the evolution of electronic laboratory notebooks includes a section about early ELNs, emphasising the importance attached at the time to providing IP protection.[42] Two articles about the RS/1 electronic laboratory notebook software appeared in the mid 1980s.[62,63] At the same time, Figueras was taking a notably optimistic view about the potential capabilities of electronic notebook systems for chemists, particularly the facilities for handling and searching structures and substructures.[64] He describes the implementation of an electronic notebook, with particular emphasis on the representation and retrieval of chemical structures, with facilities for attaching information to those structures. It is interesting to compare our 21st century expectations of electronic notebooks with the details of the system provided by Figueras.

In the mid-1990s, Myers *et al.* were extolling the virtues of electronic laboratory notebooks for collaborative research.[65] This work was to lead to the development of the Pacific Northwest National Laboratory (PNNL) notebook that we discuss in the section about Individual ELNs. In 1998, CENSA published an extensive study of the requirements for electronic notebooks as perceived at that time.[38] In the basic data capture phase of the development of ELNs, proprietary applications were much in use.[59] Several early publications described notebooks based on such proprietary software.[66–68] Other early articles reported the development of web-based notebooks.[69–71]

Requirements and capabilities

As might be expected, a number of articles have appeared as web pages; discussing the requirements, challenges, and best practices when implementing an ELN, whether commercial or bespoke. The most recent is a series of six articles under the general heading "Implementing Electronic Lab Notebooks", published under the Scientific Computing banner. The author, Bennett Lass, is the Director of ELN Services at Accelrys Inc. In this series, he examines best practices for successful ELN deployment from the following perspectives: defining and managing success; building the foundation; documenting experiments; enabling collaboration; system integration; and research management.[36]

In earlier web articles, Phillips looked at the reasons why an academic laboratory might need an ELN[72] and Elliott appraised the data and information challenges for companies.[73] Phillips bases her article on interviews with several academics and derives four reasons: generating high-throughput or automated data, collaborating with other laboratories, generating large amounts of visual data, and high personnel turnover. Her rationale rests upon the need to discover and retrieve data, either selectively or from several years previously. Phillips begins her case with a comparison:

Large companies typically use ELNs to standardize quality control or establish a legal data trail; academic labs use them to gain searchable access to years' worth of data or the ability to share data easily.[72]

Despite the apparent differences between the industrial and academic research environments, Elliott's article identifies similar issues from responses given by over 500 ELN users as part of a survey of the ELN market. The top five challenges were:

1. Finding data and information when it is needed
2. Storing and organizing data
3. Sharing data with others
4. Using too many systems and databases
5. Keeping up with the growing volume of data

We note that many if not most of these challenges are in fact very similar to the issues raised in academic research.

The University of Utah guide to ELNs provides a succinct insight into the capabilities of ELNs.[74] Williams *et al.* afford a more comprehensive evaluation of laboratory notebooks and associated data, in both hardcopy and electronic form, in an appendix to a Wiley online publication, "Current Protocols Essential Laboratory Techniques". Their extensive advice about recording experiments, from planning through to conclusions, is highly pertinent, regardless of the medium used.[34]

A recent article in *American Laboratory*, having considered the potential capabilities of multidisciplinary ELNs, concludes that modern technology offers fresh options and that *the multidisciplinary solution offers a compelling choice for organizations seeking to combine the best of the ELN world.*[75] Business requirements can be as significant as technological capabilities when implementing and deploying an ELN. In particular, the successful introduction of an ELN system can depend on matching the system workflows with those of its users.[76]

Several articles consider requirements and capabilities from the perspective of specific areas of application. Thus eCAT claims to be *the first online ELN, the first ELN to be developed in close collaboration with lab scientists, and the first ELN to be targeted at researchers in non-commercial institutions*.[77] Rubacha and Rattan put forward the five-step screening process developed at Bristol-Myers Squibb for selecting a vendor ELN[56] and other authors from the same company describe the design, validation and implementation of an ELN in their bioanalysis laboratory.[78]

A comparatively small number of articles address the challenges associated with ELN deployment specifically; typically these are conference talks for which only abstracts are available[79,80] or web articles. In one such piece, Elliott discusses change management when implementing an ELN. Drawing on surveys conducted over several years, he identifies the need to change the culture as the principal challenge. Organisations that have already implemented ELNs and those who are in the process of doing so cite resistance to change as an issue.[55] As noted earlier, the outlook for deploying ELNs in QA/QC laboratories is now more optimistic.[57]

In that QA/QC article, Metrick also draws a distinction between an ELN and a Laboratory Information Management System (LIMS), although she foresees the two becoming closer together. Wright notes that misunderstandings of terminology by laboratory staff can lead to resistance to implementation.[81] He cites the distinction between data, information, and knowledge and also that between an ELN and a LIMS: the former is intended to record what happens to a sample at a point in a workflow; the latter tracks the movement of samples through a workflow. On the positive side, Merck KGaA Darmstadt have successfully connected their ELN to a research LIMS.[82] An alternative view of the difference between an ELN and a LIMS is that the latter holds structured data whereas the former handles mostly unstructured data.[42] General acceptance of this point of view seems far from certain.

At the other end of the scale, interest in portable notebooks is growing. Macneil has written blog posts in this context. In one post, he suggested that mobile apps would change the way we do science,[83] having earlier said that apps might *overtake websites in scientific research as they now have done for general usage.*[84] Subsequently, at the end of a long post, he argued that an iPad electronic lab notebook offers *the best of both worlds*: the portability of the paper notebook and the auditability provided by electronic solutions.[27] In the same blog, he also comments on ButterflyNet, which is a mobile note taking, capture, and access system for field biologists.[85] Macneil is also a co-author of the paper describing the eCAT ELN: potential extensions include enabling *data entry and search in eCAT through smartphones.*[77]

There is ample evidence that commercial ELN providers are reaching out to tablet computing, particularly for the iPad, although in a laboratory context several approaches to the user interface issues are still being tried. The question that remains is whether the touch interface that is so attractive in the mobile consumer market can be adapted to laboratory environments.[86]

The capability to interface instruments is clearly a key requirement for most if not all ELN deployments. In 2000, when discussing the analysis of pharmaceuticals, Huynh-Ba and Aubry expressed the rather limiting opinion:

> *On a practical level, a full electronic notebook, however desirable it may be, is not practical until all the instruments in a laboratory are computerized and networked*.[87]

More recently, Hice noted that the acquisition of external data, such as from an instrument, is now essentially a case of clicking a link.[37] Security requirements are also likely to influence ELN deployments, so were the subject of a Q&A session run by Scientific Computing. This session dealt with security considerations in three question areas: verifiable electronic signatures; time stamping across multiple locations and time zones; and version control.[88]

Data and knowledge management

Taylor notes that the incorporation of data is an essential part of the R&D workflow, but (in 2005) none of the ELN vendors were achieving that effectively; he anticipated that this situation would change.[53] However, five years later, concerns remain about difficulties with exporting data from most ELNs.[89] Overall, this more recent post addresses the question of handling the research data associated with publications, very much a topical issue, but one not frequently raised in the context of ELNs. Taylor also considers that electronic laboratory environments (ELEs) still present *enormous challenges to niche suppliers of ELN systems*, owing to diversity of such environments.[42]

Kühne and Liehr, concerned that traditional approaches to information management led to loss of data or to the value of the data, proposed a new approach, relating the flow of information to the workflow of a scientist. In their introduction, they covered the retrieval and provenance disadvantages of standard filing systems. They advocate systematic information management, using a file-naming scheme based on timestamps to provide provenance (although they do not use that term).[15]

Elliott's 2005 presentation examines the role of electronic laboratory notebooks to knowledge management, comparing them to their paper equivalents. He demonstrates how ELNs contribute to knowledge sharing.[45] Four years later, the same author reports that organisations that have no specific scientific domain requirements and need only basic data capture are turning increasing to Microsoft SharePoint™.[59]

The same author has also assessed outsourced, or hosted, ELNs in the form of cloud-based services, also known as SaaS (Software as a Service). An Atrium Research survey conducted late in 2009 found: *A relatively small percentage (less than five percent) of all users currently access ELN in either a SaaS or hosted configuration.*[90]

The integration of ELNs with associated tools and with other data repositories is emerging as an area in need of attention. At the 2012 ACS meeting, Potenzone *et al*. described a search interface that federates an ELN and SharePoint.[91] With regard to tools integration, Machina and Wild propose an ELN-centric model for supporting the drug discovery and development process.[92]

Individual ELNs

Several articles describe prototype implementations of ELNs for specific laboratories: pharmaceuticals,[60] materials R&D,[93] and bioanalysis.[78] In 2002, Mackay *et al*. reported a different and appealing attempt to preserve the paper notebook paradigm in an electronic form for biologists.[94] They developed three prototypes: an A4 graphic tablet that created an online database while enabling the biologists to write as if on paper; a CrossPad version; and an augmented reality approach, the A-Book. The CrossPad prototype suffered from a lack of interactivity, while the A-Book supported interaction and linking between physical and digital objects through an interaction lens device, which Mackay *et al*. illustrate in their paper.

Mackay was also a co-author of a paper describing Prism, a hybrid paper and electronic laboratory notebook developed on the basis of observations of how biologists use laboratory notebooks.[95] They report an interesting range of practices for managing data and information in both paper and electronic forms. Prism, which they designed in participation with the users, has five key components: paper notebook, electronic notebook, desktop activity, web activity, and shared activity. Although Tabard *et al*. created Prism to be a technology probe, they say that the field-test team of bioinformaticians adapted well, organizing their activities around a master notebook.

The CombeChem e-Science project also experimented with employing a tablet PC as a capture device for chemists. This evaluation, which came to be known as the *Smart Tea* project, used an HCI (human computer interaction) approach to understand the way chemists recorded experiments, using the process of making tea as a metaphor.[96,97] *Smart Tea* was part of the overall CombeChem project concept of *SmartLab Architecture*, which included an ELN for recording experiments and the data generated by those experiments. The ELN used URIs as semantic descriptors for linking together the various entities within the *SmartLab*. The aim of this prototype architecture was *to provide effective semantic support for experimental and computational science.*[53]

ViNE, the Virtual Notebook Environment, was one of the early web-based prototypes, aiming to provide a secure collaboration environment with support for computational experiments in neurophysiology.[98] There is no evidence that work with ViNE continued after 2000.

The ELN developed at the Pacific Northwest National Laboratory (PNNL) as a collaboration and productivity tool has been in use for many years. Myers described its architecture in 2001,[99] with subsequent publications dealing with security issues[100] and semantics.[101] The semantic provisions use self-describing metadata and relationships to aggregate information generated by multiple applications and to enable browsing, searching, and reasoning across that integrated information.[102]

The Pacific Northwest team were arguably the earliest advocates of open collaboration. Kouzes, Myers, and Wulf set the scene in 1996, citing Wulf's 1993 definition of a *collaboratory* as a:

> *…center without walls, in which the nation's researchers can perform their research without regard to geographical location-interacting with colleagues, accessing instrumentation, sharing data and computational resource, and accessing information in digital libraries.*

They acknowledged the existence of both social and technological barriers to adoption, which even now we can relate to observations made previously in this paper. However, their conclusion was a confident prediction that collaboratories would be part of our future.[103] Other accounts of ELNs with collaboratory features, all of which rely to some extent on the PNNL work, are: a virtual NMR facility;[104] a biological sciences collaboratory;[105] and a geocollaboratory.[106]

Rudolphi and Goossen reasoned that the decentralized nature of academic research and its greater independence were factors contributing to the lesser use of ELNs in academia than in industry. Deeming the infrastructure requirements and cost of commercial solutions as further inhibitors to academic adoption, they designed a web-based solution using basic server software. Their open-source package, which they name *open enventory*, incorporates: a literature management system; support for experiment from the planning stage through to data collection, particularly from analytical instruments; structure and substructure search, using SMILES identifiers; and data mining functions that can extract patterns from a full range of data, subject to access permission.[107]

Many individual ELNs are now identified as brands. Atrium Research supplies a list of suppliers that *purport to develop ELN or ELN-like products*, with links to the web site for each brand.[108] One brand not listed is the eCAT ELN mentioned previously.[77] The latter paper includes an instructive functional comparison of eCAT with alternative, generic, means of recording data and information: paper, spreadsheet, database, and wiki.

Some articles feature the use of branded ELNs for specific investigations. Denny-Gouldson describes how BioBook, an IDBS product, can meet the specialized requirements of biologists.[109] Other reports describe the use of E-WorkBook Suite, also an IDBS product, in bioanalytical laboratories.[110,111]

One of the ELNs listed by Atrium Research is CyNote, now available as a cloud app from sciCloud.net, which is also interesting because its *ideological foundation of CyNote is a blog where comments can be appended onto each entry.*[112] Previously, Todoroki *et al*. had developed a blog-based notebook specifically

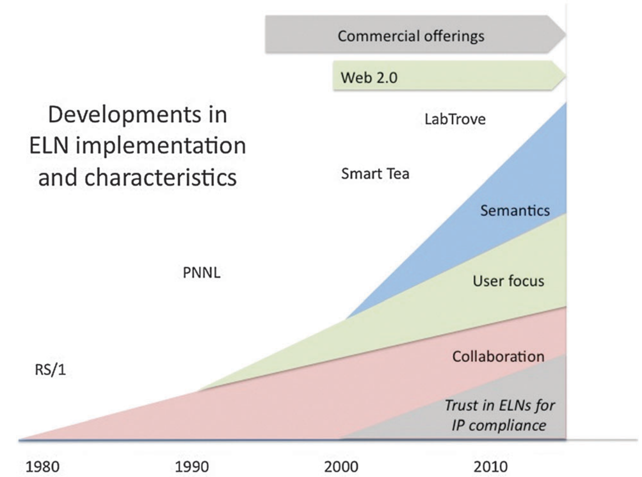

Fig. 2 The evolution of ELNs since 1980, showing the growth in support for specific characteristics.

to enable researchers to maintain a log of their activities, akin to a paper-based research notebook. They used embedded hyperlinks to access data or to refer back to an associated blog page. To control access to the blog, they added a user authentication service to the HTTP server.[113]

LabTrove is being developed at the University of Southampton, with a researcher-based frame of reference. It uses a blog-based approach to embody the journal characteristics of traditional notebooks in mind, while also incorporating the potential for linking together procedures, materials, samples, observations, data, and analysis reports. LabTrove extends the traditional blog paradigm with full access control, enabling it to meet regulatory requirements alongside flexibility for the individual user.[114]

Fig. 2 illustrates the evolution of ELNs since 1980, addressing the development of support for collaboration and semantics, the increase in user focus, and the growth in the trust placed in the ability of ELNs to satisfy regulatory requirements. We deal with regulatory issues as a specific consideration in the follow section, together with collaboration and semantics, followed by the User studies section.

For the business aspects of currently available ELNs, including comparative analyses, we refer the reader to the phaseFour Informatics website.[115]

Specific considerations

Two considerations occur routinely in discussions about the evolution and development of ELNs: knowledge protection and knowledge sharing. The need to protect knowledge arises from Regulatory and legal issues, primarily the requirement to establish intellectual property (IP) ownership. Collaboration and sharing is fundamental to science, enabling progress based on the results obtained by other scientists, subject to the test of reproducibility. To exploit knowledge, we must be able readily to access it, especially if we intend to reuse that knowledge to advance the science. The Semantic Web introduces methods and technologies that can facilitate access and reuse, so is influencing the development of ELNs, particularly in academia. In this section we look at articles dealing with these three considerations.

Regulatory and legal issues

A recent web article explains US patent law and why applications for new molecular entities tend to become litigious: the research record becomes vital to proving invention priority in a court case. *Therefore, notebook records, regardless of form, can be vital in winning or losing a patent in the U.S.*[116] The same article lists the five elements of the test for admissibility of records such as those in a laboratory notebook. The related issues of electronic record admissibility in court, together with the ability of ELNs to protect IP over long periods of time, appear in many publications about electronic notebooks. The same web article reveals a growing use of record archival services for the long-term preservation of data and authenticated information, and concludes with the advice: *ELN does not eliminate the need for proper records management*.

Attitudes to the use of ELNs are changing, as efficiency gains become more apparent and concerns about intellectual property protection are alleviated, influenced by recent patent reforms that will make "first to file" the criterion for determining the right to a patent.[117]

Vinson and Westland addressed security considerations for electronic notebooks in 1984, including the need for hardware backup, protection against malicious and inadvertent loss, access control, and meeting legal requirements. Their outlook at the time was encouraging:

The problems are not as intimidating as they might appear at first. Most of them apply with equal force to paper notebooks; those that are unique to electronic notebooks have practical solutions. Indeed, computer systems can often provide improvements over paper counterparts.[118]

The early view of the CENSA was also comparatively optimistic: *Most U.S. federal agencies have issued regulations or guidelines regarding the capture and usage of fully electronic records. For example, on March 20, 1997, the FDA issued a bonafide regulation that allows electronic records and signatures to be submitted in lieu of paper records and handwritten signatures.* However, foreseeing later concerns about the IP protection afforded by ELNs, the CENSA report gave an indication of why early adoption was slow: *None of the systems on the market today supports the legal or regulatory requirements for recordkeeping and records management well.*[38] It would not be long before writers would refer to the risks to IP posed by ELNs.[47]

Hybrid systems were one response to these concerns: organisations could print the records required for evidential purposes, authenticate them, and preserve them in traditional paper archives. One example of such a system is the BioVitrum ELN.[39,61] However, the hybrid share of the market has decreased significantly since 2006 and almost two-thirds of ELN installations are now fully electronic.[116]

By 2005, Tormey was able to cite US case law regarding the acceptance of electronic records, asserting that *electronic and paper laboratory notebooks will be treated as equivalent evidence.*[119]

He also presented a financial justification for electronic management of IP records. Nickla and Boehm, writing in the Journal of Neuroimmune Pharmacology, warn of harsh consequences if research activities are not documented properly, citing court cases and other publications. They note that the knowledge that a laboratory notebook comprises *can be critical for establishing evidence in support of intellectual property rights and for refuting claims of research misconduct.*[120] It is perhaps a sign of the times that this is one of only two publications in this survey to mention defence against claims of misconduct. The other describes the CyNote ELN:[112] both articles are recent.

Collaboration and sharing

De Roure and Frey set the scene for the contributions to collaborative knowledge acquisition that electronic laboratory notebooks can provide. Their perspectives incorporate: raw and derived data, together with its provenance; enabling reuse; and annotation in the form of rich metadata.[121] The function of electronic notebooks in collaborative research had first been put forward some ten years earlier.[38,65]

An interesting perspective on collaborative knowledge management comes from Shankar's ethnographic studies:

In the laboratory, these documentary products of daily activity represent tensions between standardization and flexibility, the collaborative nature of science and the practical and personal needs of the individual scientist, and individual learning and professional socialization. Record-keeping is more than organizational memory; in science, it has profound implications for the production of knowledge and the development of professional identity.[4]

Collaboration within an organisation can have several layers. For example, individual functions and departments can have their own procedures, which might involve differing knowledge management practices. When the nature of the business requires these departments to share information, efficiency considerations mandate a significant degree of cross-department integration is necessary, thus presenting challenges for ELN installations.[122]

Shankar also noted the natural role of annotations in processing raw data through to publications, although his proposition that formalization by annotation makes it unnecessary to store the raw data would certainly alarm any advocate of assured provenance. He quotes one researcher: *If we keep good records, we can throw away the data.*[4]

Martin *et al*. take very much the opposite view, having designed their ELN specifically to provide opportunities for users to annotate their *in silico* atmospheric chemistry experiments. They distinguish between the user-derived provenance information that relates to *why* an experiment was conducted and the system-derived provenance data that describes *how* data is generated, saying of the latter.

We view this as the extreme system-orientated perspective on provenance, completely eliminating the role of the scientist in provenance capture, which runs the risk of capturing provenance of limited value for the long-term archival of data.[123]

In their technology review, Du and Kofman included in their drawbacks section the view that improved data annotation was *essential* for ELNs *to minimize the possibility of misinterpretation, and encourage communication between users when the meaning of data needs to be clarified.*[43] Although they do not use the word *curation*, it is clear that effective curation is what they are calling for.

Macneil has discussed collaboration from an Open Research perspective and in the context of data publication in a post about the Encyclopaedia of Original Research (EOR).[89] Group openness in pre-competitive research is also relevant in this context. The Pistoia Alliance came into being in 2009, following an earlier meeting in Pistoia, Italy, with a mission to facilitate collaboration and innovation at the precompetitive stage of life sciences research. Data management and sharing issues are of particular interest, leading, amongst other topics of mutual interest, to collaborative consideration of the requirements for ELNs.[124]

For some, the apotheosis of collaboration and sharing in the ELN context is *Open Notebook Science*. The UsefulChem project, led by Jean-Claude Bradley, illustrates this concept.[125] Bradley is also a co-author of a book chapter entitled ''Collaboration Using Open Notebook Science in Academia'', ten pages of which are devoted to the history and evolution of the UsefulChem project. The final section of the chapter describes Cameron Neylon's experiences of open notebook collaboration using a blog-based laboratory notebook system developed at the University of Southampton.[126] This system subsequently evolved into the LabTrove ELN, as described at the end of the section about Individual ELNs.[114]

Dial-a-Molecule is a Grand Challenge Network that aims to generate a transformation in the speed of molecular synthesis: the vision is to make molecules in days rather than years. Open access to the results of synthesis experiments, successful and otherwise, is critical to realising this vision, as is the ability to process those records automatically.[127,128] The open publication of scientific work in this manner will not suit all researchers, and can prevent applications for patents based on that work. On the other hand, companies might still find this form of collaboration advantageous if carried out inside a firewall.[42]

Slominski notes that Web 2.0 technologies can enhance ELNs and facilitate collaboration and sharing.[129] The myExperiment approach to Open Science brings together a number of the threads in this survey, given its reasoning that researchers should share not only results and data but also methods and processes, in the form of workflows. This approach relies on collaborative data curation. The myExperiment implementation uses Semantic Web technologies.[130] The *MyExperimentalScience* project is an interesting integration of ELN and workflow that links LabTrove with myExperiment, thereby making LabTrove templates, which are preformed posts that are in effect represent single steps in a workflow, available for discovery and reuse.[131]

The Semantic Web

The most wide-ranging assessment of the current and potential impact of the Semantic Web on the recording, management, and exploitation of scientific data and information is the article entitled *The value of the Semantic Web in the laboratory.*[28] This article considers the subject from a research lifecycle perspective, appraising the impact of the Semantic Web on the planning, execution, and dissemination phases, stressing that the second phase, which requires support in the laboratory, is the most challenging.

One of the demonstrator projects for the Semantic Web in the laboratory itself was CombeChem:[53,96]

''CombeChem'' provided experience of e-science semantic support for the chemical data lifecycle, from inception in the laboratory to dissemination of data, showing how laboratory data should be recorded, using electronic laboratory notebooks, enriched with appropriate metadata, to ensure information can be correctly understood when subsequently accessed, (''Annotation@Source'').[132]

At around the same time, the PNNL electronic laboratory notebook incorporated semantic repositories and associated services, using self-describing metadata and relationships to enable browsing, searching, and reasoning across information integrated from multiple applications.[102]

Other projects that incorporate semantic enhancements and Semantic Web technologies to ELNs are: the iPad ELN; Bioclipse; and the atmospheric chemistry work.

- The iPad ELN, which is not related to the tablet device with the same name, was developed to enable biomedical researchers to record information and subsequently to associate semantic tags with that information. This approach allows for both user-generated and preset tags.[133]
- Bioclipse is an open source client workbench for chem-informatics and bioinformatics applications. It can be used to integrate data in various forms by creating semantically rich XML documents.[134]
- The atmospheric chemistry experiments discussed in the Collaboration and sharing section use Semantic Web technologies to organize the provenance information captured in the ELN.[123]

User studies

The claim of eCAT to be *the first ELN to be developed in close collaboration with lab scientists* might be a matter of interpretation.[77] There is no doubt that the developers of eCAT took full account of feedback from users of earlier online ELN prototypes, but it is open to debate whether they were indeed the first team to do so. In 2004, Achour *et al.* attributed the limited success of past attempts to develop ELNs for the pharmaceutical company Sanofi-Synthelabo to the lack of support for users. Setting aside the regulatory requirements, they developed a prototype called Kalabie based on specifications provided by their users, who were said to be *happy with Kalabie when it entered service.*[135] Agilent Technologies Inc. introduced a new version of Kalabie in June 2009.[136]

Rekik *et al.* adopted laboratory-oriented CoPs (communities of practice) as their basis for developing the eJournal notebook environment to promote collaborative research and also learning.[137,138] The Biovitrum system, described by Kihlén in 2005, was built on the basis of requirements obtained as the result of a user satisfaction survey.[61] Similarly, Tabard *et al.* developed the hybrid Prism ELN on the basis of observations of how biologists use laboratory notebooks.[95] Most recently, Ng and Ling illustrate their CyNote paper with two use cases, one of which is general note taking.[112]

Although we can see from these examples that assuming a user focus is not uncommon, accounts of formal user studies are infrequent. We have already mentioned the work of Mackay *et al.*, developing three prototype notebooks for biologists. They conducted a field study over a two-year period, in which they observed, interviewed and worked with a team of research biologists and other professionals with a stake in the recording process. They based the design of their three prototypes on five video workshops, during which they observed notebooks being updated. The summary of user requests in the paper is, as it was at the time, instructive.[94]

Shankar conducted his ethnographic study in an academic animal neuroscience laboratory over a period of eight months. His goal was not only to study record keeping in a research laboratory but also to encourage other studies of the process and its contribution to knowledge capture and management.[4]

It is a matter of record that few other information studies researchers have taken up his call, one exception being Richardson, who surveyed academic chemists from four universities to analyse their practices when recording laboratory notes and data. At the time, 51% of respondents were using paper notebooks. The charts in his Master's paper suggest that the acceptance of electronic notebooks in his four universities is markedly less than might be anticipated.[139]

The EUROCHAMP-2 project has re-engineered the ELN developed for atmospheric chemistry experiments, which emphasised the importance of provenance information.[123] The results of a questionnaire at a EUROCHAMP-2 workshop indicated that the ELN was a useful productivity aid and useful for spreading best practice.[140] A subsequent publication argued that user orientation was an important factor in tracking provenance.[141]

The most recently reported study of user practices was that of Klokmose and Zander, who interviewed PhD students and conducted workshops in the Physics and Astronomy department at Aarhus University. They used activity theory in the analysis of their results, concluding with: *There is a special need for level of flexibility which is currently absent from both desktop and web applications in general.*[142] It seems likely to the authors of this review that a significant majority of researchers and ELN users would endorse their conclusion, a view that we examine in greater depth when considering the user perspective.

Articles recounting the lessons learned about ELNs are even less common than articles about formal user studies. Manrique and Ruggles included *lessons learned in three generations of development* in a talk to the 2004 meeting of the American Chemical Society: unfortunately, only an abstract is now available.[143] Arguably, Elliott's article concerning the need to manage change when implementing an ELN is a sound generic example of a lesson learned ''the hard way''.[55]

From a different user perspective, electronic notebooks can be a useful adjunct to teaching and learning. The Collaboratory Notebook, although not strictly a laboratory notebook, enabled high school students of earth and environmental sciences to undertake task-based enquiries and to comment on each other's predictions. The Collaboratory Notebook was also used an undergraduate ecology class and in a medical school for discussing hypothetical cases.[144]

At the 2008 meeting of the American Chemical Society, Schatzberg described a research project to study the effects on first semester chemistry students of using an ELN program called LabWrite.[145] Engineering students located in different places were the subject of experiments with an electronic laboratory journal (eJournal) that enabled students to organize, analyse, and share the data that they required for their experimental assignments.[146]

The two universities in Sydney, Australia are collaborating on a project to introduce aspects of research practice into the undergraduate curriculum, in particular the use of an ELN. Beginning with selected honours and postgraduate research students, they introduced a blog-based ELN, developed at the University of Southampton, for recording thoughts, notes, observations, and data obtained from the instruments used in chemistry research.[114] After six months of their trial their students had *found this particular ELN sufficiently flexible to accommodate their experimental work*. Their plan is to extend further the use of an ELN in the undergraduate chemistry curriculum.[147–149] Iyer and Kudrle have evaluated the use of a CMS (Drupal) with ELN capabilities for an existing undergraduate course:

The initial student surveys about this approach suggest that it has been valuable in its ability to share data collectively within a given class situation.[150]

Elsewhere, at Boston University, the undergraduate organic chemistry curriculum has been remodelled to involve the routine use of an ELN and other cloud-based facilities in a teaching laboratory.[151]

The user perspective

Scientists in many disciplines regularly use paper notebooks in their work. Over the last two to three decades there has been an increasing need for workers in these disciplines to capture digital data from instruments such as mass spectrometers, geophysical equipment, or even images, sound, and video, in addition to recording paper-based notes. In the early years of computerised equipment the output was often printed out and a physical copy stuck into the paper notebook. However, nowadays such users are also likely to create digital reports in the form of word-processed documents or webpages from their notes and data for formal assessment or publication. The use of digital formats for data and reports creates a powerful incentive to capture the notes digitally using electronic laboratory research notebooks; in order to move into and stay in the digital space as early as possible in the experimental workflow.

The importance and usefulness of capturing scientific information electronically is widely recognised with the use of electronic laboratory notebooks as replacements for paper notebooks evangelised by certain parts of the scientific community, for example Elliot[45] and Mullin.[41] Unfortunately the evangelists are frequently not the same individuals that will be creating the data and recording their work in the electronic laboratory notebooks.

Difficulties arise in convincing the users who will be responsible for the data creation and information authoring in these systems of the value of the ELN because of the different requirements and expectations of for example two significant sets of users; managers and authors. The interests of the data manager or project supervisor are heavily weighted towards the future access to the data, and easy retrievability of the information, which includes some standardisation of the input. The authors of the information often do not share these same goals to the same extent, and the recording of their activities and data, and standardization of data entry,[4] is second to other aims such as performing the research activities and the interpretation of the data. Forcing users to abandon paper notebooks and instead use an ELN is unlikely to be successful, users are much more likely to adopt an ELN if they can see some proven value that is relevant to their own work.[44]

A paper notebook is not an effective tool from the point of view of the data manager or project supervisor, because it does not enable the access and sharing of data and information, it can only be read by one person at a time, it cannot be accessed anywhere, and it is difficult to search. For the authors of the content, however, the paper notebook is an incredibly easy to use, convenient, and flexible tool that allows them to focus on the task at hand, without the need for any special equipment or training. Usually, the paper notebook can be taken wherever they need it, and so they do not need to interrupt the flow of their activities in order to use it, but this is certainly not the case when controlled and restricted environments are involved. For example notebooks cannot be simply moved in and out of contained areas restricted for biohazard isolation. There are many affordances of paper notebooks, such as portability and flexibility, and not least that they use the natural techniques of writing and drawing. In comparison to this the majority of software for collecting notes is perceived to be more difficult to use, take more time, and is less flexible leading to frustration and reluctance to use the system.[152]

The ELNs are of course not without significant benefits, some from merely being ''digitally enabled''. A common benefit of ELNs is the ability to be able to link directly to digital data, for example, the embedded hyperlinks used in the blog-based electronic notebook developed by Todoroki *et al.*[113] User research carried out for the LabTrove ELN showed that both users and producers of the content recognise the value of being able to link directly to data.[152] The ELN can link to, or include, data directly from instruments or sensors, or from other ELNs, web pages, and repositories. Although the authors of content see this as a useful feature, the collation of content into a single repository is also seen as a ''simple'' way of solving the same problems, without necessarily appreciating the value of the context that linking, and other features such as metadata, provide when searching for and using the data. In 2004, Murray-Rust and Rzepa published an article challenging the transclusion model on integrity grounds.[153] They admit that their message is ''slightly tongue-in-cheek'' but go on to propose a *datument* model, in which publications contain all the relevant parts, incorporated as the datument is published.

In fact, much of the value of electronic records relies upon the content producers to do the hard work of creating the context for the record, in order for it to be findable and meaningful in the future to users other than themselves. The context for the record might be metadata describing the content, or links into other information, such as sources, background, data, and related studies. Providing accurate and meaningful context for the electronic records is required to ensure effective curation to minimize the possibility of misunderstandings about the meaning of the data. Du and Kofman raise this point in the drawbacks section of their 2007 technology review, where they indicate that improved data annotation is *essential* for ELNs *to minimize the possibility of misinterpretation, and encourage communication between users when the meaning of data needs to be clarified.*[43]

Even where institutions mandate the use of electronic repositories authors are often reluctant to make the effort to provide the metadata needed to provide context and make the content accessible by search engines.[21] Many of the content producers do not have an awareness of their potential audience or the necessary minimal training in data management that would enable them to add the most effective context or even, perhaps most significantly, have an appreciation as to why they need to do so. The lack of data management knowledge and suitable standard practices leads to minimal or worse inconsistent context being added into the electronic record.[152] For the user the activity is additional work they never needed to do before without any personal benefit.

User evaluations with ELNs for capturing atmospheric modelling data by Haji *et al.*[154] and Martin *et al.*[155] confirm that the effort involved in capturing provenance information at modelling time was likely to deter users from adopting the system. Concerns about the amount of work involved for the user in capturing the provenance information were addressed by minimising the input required through the use of automation and by prompting the user for annotation information. Evaluators were positive about the system when they were able to appreciate the value of the provenance captured in helping them to recall what took place in the experiment and its role in providing information to help other scientists and reviewers. Martin *et al.* also found that the evaluators raised concerns about whether the ELN could meet their provenance needs because the prototype did not enable the evaluators to add provenance to their model input files. Adoption of an ELN system is likely only if the system meets the needs and requirements of the authors of the data. This highlights the importance of involving the authors of the data in the design process for the ELN in order to understand their user requirements.

Attitudes to sharing

An important feature for ELNs in an environment of open data is the ability to share the notes and data with other members of the community, or even with members of the public by providing open access to the records,[156] or by providing a mechanism to publish a subset of the information. Sharing and collaboration would seem to be positive features of an ELN, but for many content producers this is seen as a feature they would not want to use, or in some cases even prevent them from using an ELN system at all. In order for data to be effectively shared it requires more detailed context and data management, including compliance with community standards for metadata and ontologies. Much more effort is required when making data available for private exchange, compared to describing it for personal use, let alone for public access.[20] This relates to many of the difficulties encountered in the early e-Science work on virtual organisations and exchanging information and services across organisational boundaries.

Willoughby found two main concerns expressed by users about using an ELN with sharing capabilities, the first around the issue of intellectual property and the second around the issue of 'readiness' of the information.[152] For the content producers the notebook represents a record of their research, together with potentially patentable or otherwise valuable material. The only people normally allowed to see the content of the notebook would be the author and their supervisor. Borgman also found that researchers had a reluctance to release their data because of concerns about issues including intellectual property, scholarly competition, misuse, and exploitation by others.[20]

For those researchers that have a concern that wider access to their data may increase the opportunities for their data to be misunderstood or misused, implementing provenance and traceability back to the source can provide attributions for quality and recognition, and can also provide evidence, if required, where mistakes or misuse have occurred.

Authors also indicate that they would not want to let anyone else see their notebook because they view it as a personal record and a 'work in progress'. Instead formal reports or scientific papers are the finished items that are meant for exposure to others. However, groups that already actively work with other groups on their research believe that being able to share the content and receive comments is a valuable feature, because they want or need to receive input from external groups to help them.

Looking ahead

There are a number of changes that need to be made for effective electronic capture of the scientific process and associated activities and data to happen. Firstly, education is needed for scientists to change the way they perceive the capture of scientific information. There is a change from science being a personal activity where the notes and data are private rough workings that are forever hidden from view, to an open process where these notes and data must be captured because they need to be found and accessed by a yet unidentified audience. This change has to provide benefits for the authors; otherwise there is no incentive to change what they do, so positive interactions have to be developed as the result of capture and sharing.

Finally, the software and hardware tools must provide the user with additional benefits to provide an incentive to use them, but they must also be as simple and efficient to use as possible. Users simply will not adopt technology that makes their lives more difficult, but you only need to look at the proliferation and success of smart phones to see that users can readily adopt and adapt to technology that provides them with capabilities that are easy to use, relevant, and useful.

Influences on the future of record keeping in the digital era

Scientific publications are in transition from the traditional reader-pays model to the open access model in which the author pays. However, it is arguable that open access publishing has little direct effect on record keeping practices. A fuller discussion is outside the remit of this paper, which focuses on preservation rather than publication. The debate about Open Access continues, but could be considered primarily an economic discussion. The debate on intelligent access to the necessary supporting data for publications[157] is in many ways much more fundamental to the pursuit of science. The disclosure of necessary information is as fundamental to a patent application as it should be for a paper. The ''need'' to minimize disclosure so as to maintain a competitive advantage in the exploitation of hard won information is understandable but is problematic for the reproducibility of the work.[158]

Nevertheless, the developing stipulations about the publication of supplementary materials will undoubtedly influence the nature of scientific record keeping in the digital era. Managing research data is a priority for JISC[22] and across the world bodies that fund research expect data to be preserved and made available; in some cases data management plans are mandatory. Journal policies for sharing research data were reviewed in 2008[159] and in the data citation guidelines published by the DCC.[24] Public policies for data management will inevitably lead to enhanced requirements concerning records management practices. Curation will assume increasing importance, with an emphasis on capturing metadata at the time data and information are created, described as curation at source.[33]

Open access and open sharing come together in Open Science, which Nielsen defined in a blog devoted to science and Web 2.0:

''Open science is the idea that scientific knowledge of all kinds should be openly shared as early as is practical in the discovery process.''[160]

In line with the Royal Society report,[157] much of the current debate now focuses not of if the information and data will be shared but when. Some scientists believe that the logical culmination of open science is to share the contents of laboratory notebooks publicly: Open Notebook Science, the subject of a recent book chapter.[126] Jean-Claude Bradley is a leading exponent of open science: he provides all the experimental results from his work on anti-malarial compounds online.[125] A co-author, Cameron Neylon, shares some of his research in a publicly accessible ELN; see, for example, his Sortase Cloning experiments.[161] Matthew Todd is another notable proponent of open notebook science, having coordinated a whole research project in public view as Project Lab Books on the ourexperiment.org site, for example the Pictet–Spengler route to Praziquantel.[156] The digital era, and the web in particular, have made it feasible to maintain a scientific record in full public view, thus drawing attention to the care required in producing that record.

Record keeping must become even more comprehensive as we advance further into the digital era. We should record failures as well as successes. For example, the success of the Dial-a-Molecule Grand Challenge will depend as much on knowing which synthetic routes do not work as on those that do.

The wide availability of online information brings responsibilities on users as well as the creators. Unless the source is known and trusted, it is always advisable to check the accuracy, and sometimes the veracity, of that source. A less obvious issue is that of copyright: students in particular might be unaware that information or data they are reusing is subject to copyright. Sometimes permission is required even to cite an item of information.

Collaboration is fundamental to the practice of science, and has become more extensive as the capabilities of the digital era have grown: indeed, open access is founded upon collaboration. In May 2008, Scientific American published an article tagged *Science* 2.0, suggesting that Web 2.0 tools such as social networking, blogging, and tagging, could facilitate collaboration and thus make science more productive.[162] Even though Crotty sounded a cautionary note two years later, drawing the distinction between talking about science and doing science,[163] there is little doubt that social networking and other 'instant' forms of communication have contributed to progress in science and technology.

We have touched upon semantics and referred to the Semantic Web throughout this paper. Their influence is most likely to bear on the nature of digital records, with the addition of semantic annotations and metadata that machines can read and reason with. Although humans can also read metadata, the likelihood is that we will become less aware of it as part of the scientific record, despite metadata being key to the future of science. It is preferable to automate the capture of metadata than to rely on humans to supply it.

We will, however, still require humans to explain their science, in accord with the stricture to ''show your working''. The Independent Climate Change E-mails Review, published in July 2010, acknowledged the need for research data, methods, and other information to be publicly accessible; for scientific debate to become more open; and for ''*all scientists to learn to communicate their work in ways that the public can access and understand; and to be open in providing the information that will enable the debate, wherever it occurs, to be conducted objectively.*''[164]

Conclusions

We have considered the nature of scientific and technical records, considering the early influences of role models such as Michael Faraday and Charles Darwin through to methods of preservation that capture the semantics of the information recorded, exploiting the technologies of the Semantic Web. ELNs are the digital equivalent of the paper notebooks kept by those role models, so we have reflected their importance in a comprehensive survey of the literature relating to ELNs. However, we recognise that it is not sufficient to provide the means of recording: we must ensure the usability of electronic recording methods. We acknowledge that the user perspective brings challenges, such as the reluctance of some researchers to share their work until they consider it complete. Some scientists find it hard to accept that failed experiments can sometimes be as informative as those that met with success.

In broad terms, scientists and science in general have gained much from the developments in electronic recording, not least for interdisciplinary and international collaboration and for the reuse and repurposing of vital data. Regrettably, the shift away from paper notebooks has brought about a diminution in the careful journaling of the thoughts and ideas leading up to scientific innovation. We are unlikely ever to see the electronic equivalent of a Faraday notebook.

Collaboration is vital for progress in all branches of science and technology, but in the digital era we do still need to increase trust in sharing. For collaboration to be effective, record keeping must become more comprehensive and provide good quality, verifiable, data and information. The means are readily available; we now need to focus on the ways.

Acknowledgements

The authors are very grateful to Fiona Bell for conducting the literature survey that forms the basis of the section entitled the electronic laboratory notebook (ELN): a literature survey and acknowledge funding from JISC for the SRF project.

References

1 P. W. Brandt-Rauf and S. I. Brandt-Rauf, *Br. J. Ind. Med.*, 1987, **44**, 68–70.
2 Wikipedia, http://en.wikipedia.org/wiki/History_of_science [Accessed 28 June 2013].
3 Library of Congress, http://www.ibiblio.org/expo/vatican.exhibit/exhibit/d-mathematics/Greek_astro.html [Accessed 28 June 2013].
4 K. Shankar, *J. Am. Soc. Inf. Sci. Technol.*, 2007, **58**, 1457–1466.
5 Wikipedia, http://en.wikipedia.org/wiki/Leonardo_da_Vinci [Accessed 28 June 2013].
6 P. Day, *The Philosopher's Tree: A Selection of Michael Faraday's Writings*, IOP Publishing Ltd, 1999.
7 J. van Wyhe, http://darwin-online.org.uk/EditorialIntroductions/vanWyhe_notebooks.html [Accessed 28 June 2013].
8 Caltech, http://www.caltech.edu/article/345 [Accessed 28 June 2013].
9 A. Eisenberg, *J. Chem. Educ.*, 1982, **59**, 1045–1046.
10 Oxford Dictionaries, http://oxforddictionaries.com/definition/english/scientific-method?q=scientific+method [Accessed 28 June 2013].
11 Merriam-Webster, http://www.merriam-webster.com/dictionary/scientific%20method [Accessed 28 June 2013].
12 C. D. Poulter, *J. Org. Chem.*, 2009, **74**, 6415.
13 C. A. Goble, J. Bhagat, S. Aleksejevs, D. Cruickshank, D. Michaelides, D. Newman, M. Borkum, S. Bechhofer, M. Roos, P. Li and D. De Roure, *Nucleic Acids Res.*, 2010, **38**, W677–W682.
14 R. L. Ackoff, *J. Appl. Syst. Anal.*, 1989, **16**, 3–9.
15 M. Kühne and A. W. Liehr, *Data Sci. J.*, 2009, **8**, 18–26.
16 C. L. Bird and J. G. Frey, *Chem. Soc. Rev.*, 2013, DOI: 10.1039/C3CS60050E.
17 D. Ince, *Significance*, 2011, **8**, 113–115.
18 M. Hulme and J. Ravetz, BBC News, 2009.
19 GreyNet_International, http://www.greynet.org/ [Accessed 28 June 2013].
20 C. Borgman, *Learn. Publ.*, 2008, **21**, 29–38.
21 J. Downing, P. Murray-Rust, A. P. Tonge, P. Morgan, H. S. Rzepa, F. Cotterill, N. Day and M. J. Harvey, *J. Chem. Inf. Model.*, 2008, **48**, 1571–1581.
22 Jisc, http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx [Accessed 28 June 2013].
23 A. Swan, *Transforming opportunities in scholarly discourse*, Jisc, Birmingham, UK, 2012.
24 A. Ball, in *Disseminate, Discover: Metadata for Effective Data Citation. DataCite, British Library, JISC Workshop*, Opus, University of Bath Online Publication Store, 2012.
25 S. M. Bachrach, *J. Cheminf.*, 2009, **1**, 2.
26 G. I. Lang and D. Botstein, *PLoS One*, 2011, **6**, e25290.
27 R. MacNeil, *The electronic lab notebook blog*. http://elnblog.axiope.com/?p=956 [Accessed 28 June 2013].
28 J. G. Frey, *Drug Discovery Today*, 2009, **14**, 552–561.
29 M. O. Jewell, E. Costanza, T. Frankland, G. Earl and L. Moreau, in *IPAW 2012 – 4th International Provenance and Annotation Workshop*, Santa Barbara, US, 2012, p. 14.
30 Digital Curation Centre, http://www.dcc.ac.uk/digital-curation/what-digital-curation [Accessed 28 June 2013].
31 W. K. Michener, J. W. Brunt, J. J. Helly, T. B. Kirchner and S. G. Stafford, *Ecol. Appl.*, 1997, **7**, 330–342.
32 C. Pancerella, J. Hewson, W. Koegler, D. Leahy, M. Lee, L. Rahn, C. Yang, J. D. Myers, B. Didier, R. McCoy, and others, in *Proceedings of the 2003 international conference on Dublin Core and metadata applications: supporting communities of discourse and practice—metadata research & applications, Dublin Core Metadata Initiative*, 2003, p. 13.
33 J. G. Frey, *International Journal of Digital Curation*, 2008, **3**, 44–62.
34 M. Williams, D. Bozyczko-Coyne, B. Dorsey and S. Larsen, in *Current Protocols Essential Laboratory Techniques*, ed. S. Gallagher and E. Wiley, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2008.
35 S. Borman, *Chem. Eng. News*, 1994, **72**, 10–20.
36 B. Lass, Implementing Electronic Lab Notebooks Part 6, *Sci. Comput.*, 2011, http://www.scientificcomputing.com/articles-IN-Implementing-Electronic-Lab-Notebooks-Part-6-102411.aspx [Accessed 9 July 2013].

37 R. C. Hice, Roadmap to a Clear Definition of ELN, *Sci. Comput.*, 2009, http://www.scientificcomputing.com/article-in-Roadmap-to-a-Clear-Definition-of-ELN-051509.aspx [Accessed 9 July 2013].

38 R. Lysakowski and L. Doyle, *Records Management Quarterly*, 1998, 23–30.

39 M. Kihlén and M. Waligorski, *Drug Discovery Today*, 2003, **8**, 1007–1009.

40 M. H. Elliott, Are ELNs Really Notebooks? *Sci. Comput. Instrum.*, 2004, http://www.atriumresearch.com/library/July%202004.pdf [Accessed 9 July 2013].

41 R. Mullin, *Chem. Eng. News*, 2003, **81**, 19–24.

42 K. T. Taylor, in *Collaborative Computational Technologies for Biomedical Research*, ed. S. Ekins, M. A. Z. Hupcey and A. J. Williams, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2011, pp. 301–320.

43 P. Du and J. Kofman, *J. Assoc. Lab. Automat.*, 2007, **12**, 157–165.

44 S. Bruce, A Look at the State of Electronic Lab Notebook Technology, *Sci. Comput.*, 2008, http://www.scientificcomputing.com/a-look-at-the-state-of-electronic.aspx [Accessed 9 July 2013].

45 M. H. Elliott, *Electronic Laboratory Notebooks: A Foundation for Scientific Knowledge Management*, Atrium Research, Wilton, CA, 2005.

46 D. Butler, *Nature*, 2005, **436**, 20–21.

47 M. C. Fitzgerald, *Chem. Innovation*, 2000, http://www.readabstracts.com/Engineering-and-manufacturing-industries/The-evolving-fully-loaded-electronic-laboratory-notebook-From-genetic-data-to-medicine-the-promising.html [Accessed 9 July 2013].

48 M. Elliott, Laboratory Informatics Guide 2008, *Sci. Comput.*, 2008, http://www.scientific-computing.com/lig2008/feature-elliot.html [Accessed 9 July 2013].

49 P. van Eikeren, *Org. Process Res. Dev.*, 2004, **8**, 1015–1023.

50 A. Nehme and R. A. Scoffin, in *Computer Applications in Pharmaceutical Research and Development*, ed. S. Ekins, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2006, pp. 209–227.

51 G. Dutton, *Genet. Eng. Biotechnol. News*, 2006, http://www.genengnews.com/gen-articles/lab-notebooks-offer-efficiency-gains/1951/ [Accessed 9 July 2013].

52 M. E. Kopach and E. A. Reiff, *Future Med. Chem.*, 2012, **4**, 1395–1398.

53 K. Taylor, J. Essex, J. Frey, H. Mills, G. Hughes and E. Zaluska, *Web Semantics: Science, Services and Agents on the World Wide Web*, 2006, **4**, 84–101.

54 R. R. Pavlis, *ELN and the Paperless Lab*, Labtronics, Inc., 2011.

55 M. H. Elliott, Change Management and ELN, *Sci. Comput.*, 2010, http://www.scientificcomputing.com/Articles-IN-Change-Management-and-ELN-061410.aspx?terms=electronic%20laboratory%20notebook [Accessed 9 July 2013].

56 M. Rubacha and A. K. Rattan, Selecting the Right ELN, *Sci. Comput.*, 2010, 4.

57 G. Metrick, QA QC: ELNs Have Come a Long Way, *Sci. Comput.*, 2011, http://www.scientificcomputing.com/articles-IN-QA-QC-ELNs-Have-Come-a-Long-Way-051111.aspx?terms=electronic%20laboratory%20notebook [Accessed 9 July 2013].

58 M. Rubacha, A. K. Rattan and S. C. Hosselet, *J. Lab. Autom.*, 2011, **16**, 90–98.

59 M. H. Elliott, What You Should Know Before Selecting an ELN, *Sci. Comput.*, 2009, http://www.scientificcomputing.com/article-in-What-You-Should-Know-Before-Selecting-an-ELN-051509.aspx [Accessed 9 July 2013].

60 C. S. Sodano, in *CINF 69*, ACS Publications, Anaheim, CA, 2004.

61 M. Kihlén, *Drug Discovery Today*, 2005, **10**, 1205–1207.

62 W. A. Gilbert, *BioScience*, 1985, **35**, 588.

63 M. F. Delaney, *J. Chem. Educ.*, 1987, **64**, 29.

64 F. John, *Graphics for Chemical Structures*, American Chemical Society, Washington, DC, 1987, vol. 341.

65 J. D. Myers, C. Fox-Dobbs, J. Laird, D. Le, D. Reich and T. Curtz, in *Proceedings of the 5th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET ICE'96)*, 1996.

66 M. Munera, R. E. Harmon, G. L. Heise and E. M. Harmon, in *Abstracts of Papers of the American Chemical Society, 195*, American Chemical Society, 1988.

67 D. Y. Pharr and F. A. Settle, *Am. Lab.*, 1996, **28**, 35.

68 G. W. Kuehnlenz, in *Abstracts of Papers of the American Chemical Society, 212*, American Chemical Society, 1996.

69 J. E. Marstaller and M. D. Zorn, *An electronic laboratory notebook based on HTML forms*, Ernest Orlando Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94720, USDOE, Washington, DC, USA, 1995.

70 J. E. Marstaller and M. D. Zorn, *An electronic laboratory notebook based on the world wide web*, Berkeley, CA 94720, 1995.

71 S. V. Hebbale and K. J. Cleetus, *Web based Electronic Laboratory Notebook Paper*, novalam.com (eBook).

72 M. L. Phillips, *The Scientist Magazine*, 2006, http://www.the-scientist.com/?articles.view/articleNo/23765/title/Do-You-Need-an-Electronic-Lab-Notebook-/ [Accessed 9 July 2013].

73 M. H. Elliott, Electronic Laboratory Notebooks Enter Mainstream Informatics, *Sci. Comput.*, 2008, http://www.scientificcomputing.com/Electronic-Laboratory-Notebooks-Enter-Mainstream-Informatics.aspx?terms=ELLIOTT [Accessed 9 July 2013].

74 University of Utah, http://campusguides.lib.utah.edu/content.php?pid=126157&sid=2131670 [Accessed 28 June 2013].

75 D. John, M. Banaszczyk and G. Weatherhead, *Am. Lab.*, 2011, 43.

76 D. J. Drake, *Drug Discovery Today*, 2007, **12**, 647–649.

77 N. H. Goddard, R. Macneil and J. Ritchie, *Automated Experimentation*, 2009, **1**, 1–7.

78 J. Zeng, M. Hillman and M. Arnold, *Bioanalysis*, 2011, **3**, 1501–1511.

79 C. J. Ruggles, in *Abstracts of Papers of the American Chemical Society, 230*, American Chemical Society, 2005, p. U1030.

80 M. P. Lee, D. Glover, P. Towers and T. Bentley, in *BIOT, 241st ACS National Meeting*, American Chemical Society, Division of Biochemical Technology, Anaheim, CA, 2011.

81 J. M. Wright, *Automated experimentation*, 2009, **1**, 1–3.

82 J. Hauss, *G.I.T. Lab. J., Eur.*, 2008, **12**, 22–23.

83 R. MacNeil, *The electronic lab notebook blog*. http://elnblog.axiope.com/?p=922 [Accessed 28 June 2013].

84 R. MacNeil, *The electronic lab notebook blog*. http://elnblog.axiope.com/?p=896 [Accessed 28 June 2013].

85 Stanford_HCI_Group, http://hci.stanford.edu/research/butterflynet/ [Accessed 28 June 2013].

86 M. H. Elliott, Tablets and ELN: A Honeymoon, *Sci. Comput.*, 2012, **29**, 9–12.

87 K. C. Huynh-Ba and A. F. Aubry, *Am. Lab.*, 2000, **32**, 13.

88 J. R. Joyce, Industry Insights: Security Considerations for Today's Electronic Lab Notebooks, *Sci. Comput.*, 2010, http://www.highbeam.com/doc/1G1-225503266.html [Accessed 9 July 2013].

89 R. MacNeil, *The electronic lab notebook blog*. http://elnblog.axiope.com/?p=1037 [Accessed 28 June 2013].

90 M. H. Elliott, ELN in the Cloud, *Sci. Comput.*, 2010, http://www.scientificcomputing.com/articles-IN-ELN-in-the-Cloud-060210.aspx?terms=electronic%20laboratory%20notebook [Accessed 9 July 2013].

91 R. Potenzone, P. McHale, M. Schoenberg, A. Jewett, K. Blanchard, D. Gosalvez and P. Skinner, in *Abstracts of Papers, 244th ACS National Meeting & Exposition*, American Chemical Society, Philadelphia, PA, 2012.

92 H. K. Machina and D. J. Wild, *J. Lab. Autom.*, 2012, **8**, http://jla.sagepub.com/content/18/2/126 [Accessed 9 July 2013].

93 S. M. Thornberg, in *Sixth Forum for Laboratory Informatics*, Philadelphia, PA, 2010.

94 W. E. Mackay, G. Pothier, C. Letondal, K. Bøegh and H. E. Sørensen, in *Proceedings of the 15th annual ACM symposium on User interface software and technology*, ACM, Paris, France, 2002, vol. 4, pp. 41–50.

95 A. Tabard and W. Mackay, in *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, 2008, pp. 569–578.

96 G. Hughes, H. Mills, D. De Roure, J. G. Frey, L. Moreau, M. C. Schraefel, G. Smith and E. Zaluska, *Org. Biomol. Chem.*, 2004, **2**, 3284–3293.

97 M. C. schraefel, G. Hughes, H. Mills, G. Smith, T. Payne and J. Frey, in *CHI 2004*, ACM Press, 2004.

98 A. D. Malony, J. E. Cuny, J. L. Skidmore and M. J. Sottile, *Future Generation Computer Systems*, 2000, **16**, 453–464.

99 J. D. Myers, E. S. Mendoza and B. L. Hoopes, in *Proceedings of the IASTED International Conference, Internet and Multimedia Systems and Applications*, ed. M. H. Hamza, IASTED, Acta Press, Anaheim, CA, 2001, pp. 334–338.

100 J. D. Myers, in *Proc. CTS'03*, Society for Modeling and Simulation International, San Diego, CA, Orlando, Florida, 2003.

101 T. Talbott, M. Peterson, J. Schwidder and J. D. Myers, in *Proceedings of the 2005 international conference on Collaborative technologies and systems*, 2005, pp. 136–143.

102 J. D. Myers, M. Peterson, K. P. Saripalli and T. Talbott, in *Abstracts of Papers of the American Chemical Society, 227*, American Chemical Society, 2004, p. 689.

103 R. T. Kouzes, J. D. Myers and W. A. Wulf, *Computer*, 1996, **29**, 40–46.

104 K. A. Keating, J. D. Myers, J. G. Pelton, R. A. Bair, D. E. Wemmer and P. D. Ellis, *J. Magn. Reson.*, 2000, **143**, 172–183.

105 G. J. Chin and C. S. Lansing, in *CSCW '04 Proceedings of the 2004 ACM conference on Computer supported cooperative work*, ACM, New York, NY, USA, 2004, vol. 6, pp. 409–418.

106 A. M. MacEachren, W. Pike, C. Yu, I. Brewer, M. Gahegan, S. D. Weaver and B. Yarnal, *Computers, Environment and Urban Systems*, 2006, **30**, 201–225.

107 F. Rudolphi and L. J. Goossen, *J. Chem. Inf. Model.*, 2012, **52**, 293–301.

108 Atrium_Research, http://www.atriumresearch.com/html/eln.htm [Accessed 29 June 2013].

109 P. Denny-Gouldson, *Am. Lab.*, 2007, http://www.americanlaboratory.com/914-Application-Notes/35149-Meeting-the-Specialized-Requirements-of-the-Biologist-Through-a-Biology-Focused-Electronic-Laboratory-Notebook/ [Accessed 9 July 2013].

110 B. Beato, A. Pisek, J. White, T. Grever, B. Engel, M. Pugh, M. Schneider, B. Carel, L. Branstrator and R. Shoup, *Bioanalysis*, 2011, **3**, 1457–1470.

111 J. Rajarao and S. Weiss, *Bioanalysis*, 2011, **3**, 1513–1519.

112 Y. Ng and M. H. T. Ling, in *The Python Papers Monograph 2:15. Proceedings of PyCon Asia-Pacific*, 2010, pp. 1–13.

113 S. Todoroki, T. Konishi and S. Inoue, *Appl. Surf. Sci.*, 2006, **252**, 2640–2645.

114 A. J. Milsted, J. R. Hale, J. G. Frey and C. Neylon, *PLoS One*, 2013, submitted.

115 J. Trigg, http://www.phasefour-informatics.com/ [Accessed 29 June 2013].

116 M. H. Elliott, New Debates over Intellectual Property Protection and ELN, *Sci. Comput.*, 2011, http://www.scientificcomputing.com/articles-IN-New-Debates-over-Intellectual-Property-Protection-and-ELN-010711.aspx?terms=electronic%20laboratory%20notebook [Accessed 9 July 2013].

117 M. Swartz, P. Skinner, P. McHale, A. Jewett and K. Blanchard, in *Abstracts of Papers, 244th ACS National Meeting & Exposition*, American Chemical Society, Philadelphia, PA, 2012.

118 J. W. Vinson and R. D. Westland, in *Abstracts of the Papers of the American Chemical Society, 188*, American Chemical Society, 1984.

119 P. Tormey, *Bio-IT World*, 2005.

120 J. T. Nickla and M. B. Boehm, *J. Neuroimmune Pharmacol.*, 2011, **6**, 4–9.

121 D. De Roure and J. Frey, in *Workshop on Semantic Web for Collaborative Knowledge Acquisition (SWeCKa)*, Hyderabad, India, 2007, p. 7.

122 M. H. Elliott, Thinking Beyond ELN, *Sci. Comput.*, 2009, http://www.scientificcomputing.com/articles-IN-Thinking-Beyond-ELN-120809.aspx?terms=electronic%20laboratory%20notebook [Accessed 9 July 2013].

123 C. J. Martin, M. H. Haji, P. K. Jimack, M. J. Pilling and P. M. Dew, *Philos. Trans. R. Soc., A*, 2009, **367**, 2753–2770.

124 Pistoia_Alliance, http://www.pistoiaalliance.org/ [Accessed 29 June 2013].

125 UsefulChem_Wiki, http://usefulchem.blogspot.fr/ [Accessed 29 June 2013].

126 J.-C. Bradley, A. S. I. D. Lang, S. Koch and C. Neylon, in *Collaborative Computational Technologies for Biomedical Research*, ed. S. Ekins, M. A. Z. Hupcey and A. J. Williams, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2011.

127 S. Coles, R. Whitby, J. Frey, C. Bird and A. Day, in *Abstracts of Papers, 244th ACS National Meeting & Exposition*, Philadelphia, PA, 2012.

128 S. Coles, G. Tizzard, J. Frey, A. Milsted, M. Edwards, R. Onyeabo, J. Spencer and J. Kuras, in *Abstracts of Papers, 244th ACS National Meeting & Exposition*, Philadelphia, PA, 2012.

129 A. A. Slominski, in *Proceedings of the 2010 6th World Congress on Services*, 2010, pp. 326–327.

130 D. De Roure, C. Goble, S. Aleksejevs, S. Bechhofer, J. Bhagat, D. Cruickshank, P. Fisher, D. Hull, D. Michaelides, D. Newman, R. Procter, Y. Lin and M. Poschen, *Concurrency and Computation: Practice and Experience*, 2010, **22**, 2335–2353.

131 J. G. Frey, A. Milsted, D. Michaelides and D. De Roure, *Concurrency and Computation: Practice and Experience*, 2012, **25**, DOI: 10.1002/cpe.2922.

132 J. G. Frey, in *Abstracts of Papers of the American Chemical Society, 232, CINF 5*, American Chemical Society, 2006.

133 A. Polonsky, A. Six, M. Kotelnikov and V. Polonsky, in *ESWC'06 Industry Forum: How Business application challenges Semantic Web Research*, ed. A. Leger, A. Kulas, L. Nixon and R. Meersman, Budva, Montenegro, 2006.

134 T. Helmus, S. Kuhn, P. Murray-Rust, M. R. Cherto, H. S. Rzepa, O. Spjuth, C. Steinbeck, J. E. S. Wikberg and E. Willighagen, in *CINF 95 – Abstracts of Papers of the American Chemical Society, 233*, 2007.

135 Z. Achour, T. Laidboeur, O. Gien, A. Musolino, X. Bon and B. Grimaud, *Org. Process Res. Dev.*, 2004, **8**, 983–997.

136 Agilent Technologies, http://www.agilent.com/about/newsroom/presrel/2009/23jun-ca09047.html [Accessed 29 June 2013].

137 Y. Rekik, D. Gillet, A. V. Nguyen-Ngoc and T. Guillaume-Gentil, in *7th International Conference on Information Technology Based Higher Education and Training*, Sydney, Australia, 2006.

138 Y. Rekik, D. Gillet, S. E. L. Helou and C. Salzmann, *Int. J. Web Base. Learn. Teach. Tech.*, 2007, **2**, 61–76.

139 K. A. Richardson, North Carolina at Chapel Hill, 2009.

140 M. H. Haji, Z. M. Zaki, P. Dew, L. Lau, J. Young, A. Rickard and M. Pilling, in *UK e-Science All Hands Meeting*, 2010, p. 3.

141 Z. M. Zaki, P. M. Dew, M. H. Haji, A. Rickard and J. Young, in *IEEE Seventh International Conference on eScience*, 2011, pp. 371–378.

142 C. N. Klokmose and P. Zander, in *Proceedings of COOP 2010, Computer Supported Cooperative Work*, ed. M. Lewkowicz, P. Hassanaly, V. Wulf and M. Rohde, Springer London, London, 2010, pp. 119–140.

143 J. Manrique and C. J. Ruggles, in *Abstracts of Papers of the American Chemical Society, 227*, American Chemical Society, 2004.

144 D. C. Edelson, R. D. Pea and L. M. Gomez, *Commun. ACM*, 1996, **39**, 32–33.

145 W. E. Schatzberg, K. A. O. Pacheco and J. P. Suits, in *CHED 1441-Abstracts of Papers of the American Chemical Society, 235*, American Chemical Society, 2008.

146 G. J. Fakas, A. V. Nguyen and D. Gillet, *Computer Supported Cooperative Work (CSCW)*, 2005, **14**, 189–216.

147 R. Quinnell, D. B. Hibbert and A. Milsted, in *Proceedings ascilite*, Auckland, 2009, pp. 799–803.

148 R. Quinnell and D. B. Hibbert, in *International Conference on Education, Training and Informatics: ICETI*, Orlando, Florida, 2010.

149 B. Hibbert, J. G. Frey, R. Quinnell, M. Mocerino, M. Todd, P. Niamsup, A. Plummer and A. Milsted, in *Proceedings of the 16th UniServe Science Annual Conference*, Sydney, 2010, p. 124.

150 R. Iyer and W. Kudrle, *Technol. Interface Int. J.*, 2012, **12**, 5–12.

151 S. P. Mulcahy, J. A. Bishop, S. E. Schaus and J. K. Snyder, in *Abstracts of Papers, 243rd ACS National Meeting & Exposition*, American Chemical Society, San Diego, 2012.

152 C. Willoughby and J. G. Frey, *Private Communication*, 2013.

153 P. Murray-Rust and H. S. Rzepa, *J. Digital Inf.*, 2004, **5**, http://www.ch.ic.ac.uk/rzepa/jodi/ [Accessed 9 July 2013].

154 M. H. Haji, P. M. Dew and C. Martin, in *2008 Microsoft eScience Workshop*, 2008, pp. 135–137.

155 C. Martin, M. H. Haji, P. Dew, M. Pilling and P. Jimack, in *Provenance and Annotation of Data and Processes*, 2008, pp. 293–308.

156 M. Todd and *et al.*, http://www.ourexperiment.org/racemic_pzq [Accessed 29 June 2013].

157 G. Boulton, P. Campbell, B. Collins, P. Elias, W. Hall, G. Laurie, O. O'Neill, M. Rawlins, J. Thornton, P. Vallance and M. Walport, *Science as an open enterprise*, The Royal Society, 2012.

158 Academic_Evolution, http://www.academicevolution.com/2009/05/galileo-open-access.html [Accessed 29 June 2013].

159 H. A. Piwowar, W. W. Chapman and W. Chapman, in *AMIA … Annual Symposium proceedings/AMIA Symposium. AMIA Symposium*, 2008, pp. 596–600.

160 M. Nielsen, http://www.openscience.org/blog/?p=454 [Accessed 29 June 2013].

161 C. Neylon, http://blogs.chem.soton.ac.uk/sortase_cloning [Accessed 29 June 2013].

162 M. M. Wardrop, *Scientific American. Science 2.0 – Is Open Access Science the Future?*, 2008.

163 D. Crotty, *The Scholarly Kitchen*. http://scholarlykitchen.sspnet.org/2010/02/08/, 2010.

164 M. Hulme and J. Ravetz, http://news.bbc.co.uk/1/hi/sci/tech/8388485.stm [Accessed 29 June 2013].