Article Type: Comparative Analysis Review

Considerations for Implementing Electronic Laboratory Notebooks in an Academic Research Environment

Stuart G. Higgins[1,2,3], Akemi A. Nogiwa-Valdez[1,2,3], Molly M. Stevens[1,2,3]*

*email: m.stevens@imperial.ac.uk

[1]Department of Materials, Imperial College London, London, SW7 2AZ, UK.

[2]Department of Bioengineering, Imperial College London, London, SW7 2AZ, UK.

[3]Institute of Biomedical Engineering, Imperial College London, London, SW7 2AZ, UK.

Abstract

As research becomes predominantly digitalised, scientists have the option of using electronic laboratory notebooks to record and access entries. These systems can more readily meet volume, complexity, accessibility and preservation requirements than paper notebooks. Whilst the technology can yield many benefits these can only be realised by choosing a system that properly fulfils the requirements of a given context. This review explores the factors that should be considered when introducing electronic laboratory notebooks to an academically focused research group. We cite pertinent studies and discuss our own experience implementing a system within a multi-disciplinary research environment. We also consider how the required financial and time investment is shared between individuals and institutions. Finally, we discuss how electronic laboratory notebooks fit into the broader context of research data management. This article is not a product review; it provides a framework for both the initial consideration of an electronic laboratory notebook and the evaluation of specific software packages.

Main

Electronic Laboratory Notebooks are a tool to efficiently capture experimental metadata

Alongside millions of new research publications each year,[1] is the creation of millions more laboratory notebook entries. These contain important metadata, reflecting the nuance of experimental work. The ability to readily access, use and share laboratory notebook data allows researchers to quickly infer meaning from results and can help facilitate reproducibility across experiments. Collaborative or multi-disciplinary research fields require efficient methods for capturing and sharing notebook entries between a diverse range of scientists.

Research relies upon computing to analyse and present data, therefore storing laboratory notebook entries in a digital format allows them to sit seamlessly alongside research data as it is processed. Electronic laboratory notebooks (ELNs) are fundamentally a means of digitising entries at the point of creation, enabling that data to be processed computationally. However, they are not a panacea. Before deploying an ELN, it is critical that the requirements of users, as well as the advantages and disadvantages of different approaches, are properly understood to avoid creating a system that hinders rather than helps.

The past 20 years have seen a rapid increase in the number of Electronic Laboratory Notebook software packages

ELNs have been mooted in various forms since the late 1950s.[2] In the 1980s, software such as RS/1 (Bolt, Beranek and Newman, Inc.) offered researchers the capability to store, analyse and comment upon data.[3,4] ELNs are presented as a tool for improving the reproducibility of research by facilitating the transfer of vital experimental details, both between generations of researchers and across different research groups.[5,6] Recording, accessing and preserving paper-based records can be slow, inefficient and difficult to integrate with modern computer-controlled data capture systems. However, implementing an

ELN is non-trivial. Its adoption requires clear understanding of notebook use in a given laboratory setting, and the provision of sufficient resources.

The majority of current ELNs are commercial offerings. These offer access to proprietary software, typically hosted remotely and available via subscription, under a software as a service (SaaS) business model. A few community-developed open-source ELNs exist, with freely accessible codebases. There are also a small number of commercial ELNs with open-source codebases, and free (to non-profit organisations) ELNs with proprietary codebases. Reviewing specific products is beyond the scope of this article, however, a number of product comparisons are available online.[7–10]

In the past twenty years, the number of ELNs has increased dramatically, as the benefits of digitisation have been recognised (**Figure 1**). In this marketplace not all ELNs have proven successful. A significant number of both commercially available and open-source software packages have ultimately become defunct. In our analysis (**Supplementary Method 1**) of 172 ELN products (96 active, 76 defunct), the average lifetime of an ELN was found to be 7 ($\pm$ 4.4) years (median $\pm$ consistent median absolute deviation). The lifetimes breakdown as 6 ($\pm$ 4.4) years (n = 25) and 7 ($\pm$ 4.4) years (n = 147), for open-source and proprietary codebases respectively. The longest running open-source ELN in our survey was ELOG (Stefan Ritt, Paul Scherrer Institut),[11] which has been active for 20 years. The longest running proprietary ELN we found was Gene Inspector (Textco BioSoftware, Inc.),[12] which has been active for 25 years. Company acquisitions, changes in the commercial market and lack of developer support or funding for open-source projects, can all result in defunct ELNs. Long-term support and data access should be a primary concern when implementing an ELN: many benefits (for example, rapid access to historic notebook entries) disappear if archived material is trapped inside inaccessible legacy systems, or worse, deleted. Procedures for extracting and archiving data in accessible formats should be part of any deployment strategy.

Before choosing an ELN the purpose of the laboratory notebook must be identified

A laboratory notebook serves a variety of purposes. For the researcher it is a record of experiments and work conducted. The notebook may describe experimental methods, be a direct record of original data, or provide metadata required to contextualise other data. Formal metadata (experimental test parameters and control conditions) may be supplemented by unplanned observations and annotations, both facilitating data analysis and interpretation. A notebook may journal both the genesis of ideas and the decision making process.[18] Successfully capturing this information is critical to the researcher and others attempting to replicate the work.

Laboratory notebooks can be used to enforce good practice and to standardise workflows. For example, institutions may mandate the inclusion of risk assessment templates within synthetic chemistry notebooks to encourage researchers to identify and mitigate hazards immediately prior to conducting a reaction. Routine procedures with well-defined outputs may follow a standard notebook template, to streamline information capture and standardise record keeping, or to record quality control procedures. This process can aid adherence to best research practices, such as ensuring enough details are captured to facilitate reproducibility. An ELN can simplify this by acting as a database of templates and protocols.[16,19] For researchers, managers and institutions, laboratory notebooks provide evidence of work completed, facilitating internal accountability, and providing a legal record to demonstrate regulatory compliance and potentially aiding intellectual property protection.

Identifying a given laboratory's requirements defines and constrains the choice of ELN. Academic research laboratories typically feature a diverse range of experiments, data types and disciplines, resulting in users having different requirements from the same ELN package.[20] While many ELNs are specialist products, targeting researchers in a specific domain (e.g. biochemistry or pharmacology), these may not be relevant or sufficiently flexible for most researchers. Whilst this limitation was recognised by the early creators of ELNs in the 1990s,[21] it remains an issue that has been highlighted again in recent studies.[16] Record keeping may be required to meet regulatory standards, for example for laboratories accredited to the testing and calibration standards ISO 17025:2017 and ISO 15189:2012,[22,23] which stipulate requirements for laboratory information management, or for those wishing to adhere to general electronic record keeping standards such as the Code of Federal Regulations Title 21 Part 11.[24] The motivating factors and requirements for a research group may not be the same as those of other groups at the same host institution, so care must be taken to identify the priorities of different stakeholders before selecting a particular product.

ELNs recover researcher time and enable better research practices, in return for financial cost

ELNs provide quality-of-life improvements over paper notebooks. In environments where ELNs have been implemented, the ease with which information can be sought and shared is regarded as one of the key benefits.[19,25] **Figure 2** illustrates different ways information can be shared via an ELN. Making ELN entries visible to multiple users is often simple to accomplish within the software. Project teams can instantly access relevant experimental data from different researchers, facilitating project management. Supervisors can remotely provide feedback without physical access to a notebook and add digital signatures to verify entries. Collaborators can be geographically separate, working in separate laboratories across multiple countries. Similarly, researchers can access their records from multiple locations, for example from different laboratories or from home. This can be an advantage where physical access to facilities is restricted, as seen during the COVID-19 pandemic. It also mitigates the need to transport physical laboratory notebooks between locations, reducing the risk of cross-contamination and data loss.

Notebook entries can be archived in situ as team members leave, while allowing incoming students and staff instant access. The ability to rapidly search through all available content allows researchers to filter and access ELN entries pertinent to their own work.[18,25] When users leave an organisation, if permitted, an ELN allows them to create copies of their entries for future reference. This can allow quick access to previous and ongoing research ideas as a researcher progresses through their career. Physically storing and preserving digital data over long periods is more space and time efficient than attempting to store paper notebooks.

Table 1 illustrates how the differences between interacting with paper and electronic laboratory notebooks are ultimately a choice between time and money. Paper notebooks can be an inexpensive medium but executing actions that are trivial with an ELN (searching, sharing, data backup, etc.) are time expensive. Conversely, most ELNs are relatively expensive to implement and maintain, compared to paper notebooks, but provide far greater functionality at a much lower time cost to the researcher. Actions such as searching, reordering, sharing and archiving can be extremely fast compared to paper notebooks. Depending on the implementation, ELNs may introduce a time delay for users as they require turning on hardware and authenticating into the software before a notebook entry can be made. The degree to which this is an issue depends on the hardware (a modern tablet can wake and unlock within a second) and software (biometric authentication can offer rapid logins, or, more commonly, local policy can dictate how long user sessions remain active before forcing reauthentication).

The operating environment impacts both paper and electronic notebooks. Laboratories that contain some form of protected environment (for example cleanrooms or biological containment laboratories) may have restrictions on the movement of items into and out of the space. Both paper notebooks and computer hardware can be contaminated in the laboratory. An ELN may help alleviate these issues by allowing access to notebooks via devices that remain inside the protected environment. However this requires pre-planning of hardware requirements.

When considering any ELN system the benefits of enabling more time to be dedicated to research, improving knowledge transfer and experimental reproducibility, are balanced against financial costs. Recognising who will bear these costs is important. For example, relying upon users to provide their own computing device to access an ELN effectively transfers this cost onto the researcher. This may cause people to spurn ELN adoption and disadvantages researchers without existing devices, as seen in studies of undergraduate

web-based learning technologies.[26,27] If a laboratory already has a sufficient number of network connected workstations, then this cost may be nil, otherwise it can form a significant proportion of the overall implementation cost (a factor that is not included in any software vendor pricing).

Interaction	Paper notebook	Electronic notebook
Adding entries to notebook	Trivial in most environments	Initial delay to sign in to system, then rapid Physically typing entries can be challenging in certain laboratory settings (e.g. when wearing personal protective equipment)
Embedding hand-drawn illustrations, equations, chemical structures	Trivial, can be drawn directly into notebook	May be challenging without appropriate hardware (e.g. a stylus) and compatible ELN. Some ELNs incorporate specialist equation/structure editors
Embedding non-text computer generated content (e.g. digital photographs, screenshots, code snippets)	Requires printing, followed by pasting into notebook	Trivial; may be embedded in-line with text or attached as file
Searching through entries	Time consuming, reliant upon user constructed index	Fast, searchable by keyword, time, user, metadata tags, etc
Reordering entries	Impossible, limited to chronological order, challenging when undertaking multiple concurrent experiments	Dynamic, instantly able to sort and present entries according to metadata
Using templates, linking to common equipment/reagent information, or standard operating procedures (SOPs)	Printing and pasting documents into notebook, or reliant on written reference – time consuming and static	Templates, equipment/reagent details, or SOPs can be held in a central database and dynamically linked directly to one or more entries
Reading entries (intelligibility, accessibility)	Reliant on user to write clear, intelligible notes	Standardises input as typewritten text
	Limited accessibility	Potential to improve accessibility using handwriting and/or voice recognition software

Maintaining a legal record	Facilitated by signing and co-signing notebook pages, plus secure storage of notebook	Facilitated by digital signatures/timestamping and robust backup regime
Sharing notebook entries	Time intensive, involving scanning and sending relevant entries, or run risk of notebook loss	Trivial; a hierarchy of permissions can be established for a given entry Data access can be provided instantly, worldwide if required
Backing up notebooks	Time intensive to scan entire notebook	Trivial – data can automatically be copied to multiple locations on user-defined basis
Rectifying accidental damage to notebook	Inexpensive to replace notebook	Expensive to replace computing hardware
	Data may be irrecoverable	For client-server architectures, data is preserved separate to device
Physically storing notebooks	Over time can amount to significant space requirements for larger laboratories	Institution-level storage possible on single server, with possibility to outsource storage to data centre
Accessing archived or historical entries	Time intensive	An ELN with an appropriate data export function can facilitate building rapidly accessible/searchable digital archive
Using notebook in a controlled environment (e.g. cleanrooms, containment laboratories)	Specialist notebooks may be required, notebook must remain within laboratory	ELN can provide seamless access inside and outside of controlled environment, assuming suitable access devices provided
Interoperating with other laboratory systems (e.g. measurement /inventory/procurement databases)	Manual static reference	Possibility to incorporate automated/direct links using persistent identifiers and APIs, dependent on ELN support
Sharing entries as open research data/open science	Time intensive as per other sharing issues	Built-in support with some ELNs

Table 1: Different laboratory notebook interactions, comparing their ease between paper and electronic media.

An ELN is typically a combination of user-interface, centralised database and file store

Figure 3 illustrates a simplified view of a commonly adopted ELN architecture (although many different approaches are possible).[11,13,28–30] Notebook entries may be stored in a relational database with attached files in a file store, with the ELN software facilitating user access and defining how notebook entries can be written and read. The software may be accessed via a web browser or in some cases through a custom application written for a specific platform (e.g. desktop or mobile operating system apps). Application-based ELNs may cause compatibility issues in academic research environments that typically feature a diverse range of operating systems.[31] Depending on the implementation, the separate ELN components may be separate servers,[30] in different physical locations. This major conceptual shift from paper notebooks brings both opportunities and challenges. While the underlying technology is ideally invisible to the end user, the choice of ELN can influence the availability of different features. For example, most ELNs are ill-suited to storing large volumes of raw data (either from a performance or cost perspective). A locally-hosted ELN server may rapidly run out of physical storage space if not appropriately provisioned. A cloud-hosted server may quickly incur significant hosting costs as storage and bandwidth requirements increase. Even with these restrictions, the amount of information that can be stored is more than is possible in paper notebooks, which can only store small quantities of data.

Sensitive data, for example patient or commercially-sensitive information, may fall under local institution or legally-mandated data protection regulations, for example GDPR 2016/679,[32] which dictates the handling and safeguarding of personal data within the European Union. This may restrict the physical location and transfer of data, excluding the use of off-premises ELNs that use internationally-based cloud-based infrastructure.[33] Locally-hosted ELNs may provide greater control over data security by restricting notebook access to users inside an institutional network (i.e. access devices must be physically on premises or connected to the internal network remotely, for example using a virtual private network).

Data integrity can be enhanced using version control and timestamping

The ability to create immutable notebook entries which cannot be removed or altered after creation is critical for academic and legal integrity. Paper notebooks typically implement this through rules and procedures, for example, by prohibiting the removal of pages, and signing and counter-signing entries. Not all ELNs address this issue, again precluding the use of some products. For example, general note-taking software packages, such as Microsoft OneNote and Evernote, do not typically include features to digitally sign or timestamp entries, with workarounds such as signing exported files required.[31,34] The level of verification required depends on both regulatory requirements and locally accepted laboratory notebook standards. For example, it may be sufficient to rely upon administrator-defined software features, such as the ability of a supervisor to lock notebook entries to prevent them from being edited or to disable entry deletion. Many ELN packages include a mechanism for version tracking, which record how a document is edited over time, a potential deterrent to modifying entries after the fact. This is a conceptual difference between paper and electronic notebooks. In a paper notebook, entries are instantly recorded (presuming some form of indelible ink is used). In an electronic notebook, there is typically a finite amount of time during which the entered text or other content remains malleable before it is saved to the server. This period can range from seconds to minutes, with the server creating intermediate versions of the entry, or a longer period until the entry is locked or finalised through some technical means. Excessive versioning may significantly increase

storage requirements and overall running costs, hence local policy is required to determine a suitable compromise for the period of time between making an entry and some form of versioning and ultimately finalisation.

If verifying the provenance of notebook entries is critical, the ELN should incorporate technologies that adhere to recognised standards for trusted timestamping, such as RFC 3161,[35] or ANSI X9.95-2016.[36] Trusted timestamping uses an audited third-party organisation (typically a commercial provider), to digitally sign and timestamp a file. A cryptographic hashing algorithm is applied to a digital file, for example a PDF file that corresponds to a notebook entry. The algorithm generates a file containing a hash (a mathematically unique representation of the original data). This hash cannot easily be reverse engineered to recreate the original data so can be safely transmitted to the trusted timestamping authority via the internet. The authority digitally signs and timestamps the hash, in the process generating a third file (a timestamp token). This token is sent back to the ELN software and stored alongside the original PDF file. Any modifications to the original file will invalidate the token, as recalculating the hash of the modified file will result in one different to the value contained within the digitally signed token. The process allows a digital file's contents at a given point in time to be verified. Although technically complex, this process is typically performed invisibly to the user.

While trusted timestamping incorporates the concept of digital signatures, it adds the additional benefit of not only verifying the signer's identity, but also the time at which an entry was timestamped. In practical terms, local policy is required to ensure users routinely timestamp their notebook entries, as the date and time of timestamping is being recorded, not the instant when an experiment was conducted. Procedures for archiving data should ensure both the timestamped file and timestamp token are properly preserved.

Open-source and commercial electronic laboratory notebooks have different costs and benefits

In our survey, commercial ELNs were observed to be far more prevalent than open-source ELNs (147 vs 24 products identified). **Table 2** compares commercial to open-source approaches. While SaaS is a widely adopted licensing model (for example, institutional subscriptions to Microsoft Office 365 or Google Workspace), SaaS ELNs may be prohibitively expensive for individual research groups due to per-user pricing, ever-growing file storage costs, and an indefinite subscription required to maintain access. In 2017, Kanza et al. reported that, in a survey of 169 users participating in an ELN pilot study, both limited financial budgets and the time required to implement an ELN were major concerns.[16] Similarly, while open-source ELNs can have relatively low initial and ongoing financial costs, they require time to run and maintain, and may require server hardware to be purchased. The relatively modest requirements for many open-source ELNs make it feasible to repurpose old computing hardware to act as an ELN server, helping to reduce this cost. Costs can scale with the size of deployment, for example commercial site-wide licenses may be negotiated at preferential rates than the per-user pricing available to individual research groups.[37] Implementing an open-source ELN at institutional scale can take advantage of pre-existing server infrastructure and ICT support. Some commercial providers offer free or reduced pricing for academic users.[16]

Open-source ELNs have the benefit of allowing not only the data to be archived, but also the underlying software itself. Technologies such as virtualisation and containerisation present feasible pathways to preserve the operating environment of the ELN for future access, beyond software and hardware obsolescence. For example, provided the codebase has been properly archived, open-source ELN software may be resurrected as a virtual machine

within a modern operating system, to allow historic file access and export. This can help facilitate future file interoperability by providing easy access to the original software. The open codebase also means data formats and standards are fully exposed, facilitating the future development of tools to reparse or extract data. When assessing an ELN hosted by a third-party, consider what will happen to notebook entries when the product is discontinued. Some open-source projects (such as the Open Science Framework)[38] may include contingency plans and funds to ensure the ongoing preservation of research data. If an ELN provides data export functions, these should be tested to confirm they provide the required level of functionality. For commercial ELNs, it should be ensured that supplier contracts include the necessary terms to facilitate data export.

Feature	Open source	Commercial
Ease of deployment	Requires client to have knowledge and/or hardware to implement	Available 'out-of-the box'
Support available	Online contributor community / may or may not have limited developer support	Company dependent, may offer training sessions, webinars, etc
Data backup	Self-arranged – flexible and full control, requires ongoing monitoring and verification	Either automatic by the vendor (e.g. automated cloud-based backups) or self-arranged
Initial costs	No software cost	No, or small, initial software cost
	Server hardware costs (and/or private cloud-server rental)	No server hardware cost (if cloud hosted by supplier)
	Provision of access devices to users	Provision of access devices to users
	Time cost for user training and migration	Time cost for user training and migration
	Time cost to setup and integrate with existing institutional facilities	Time cost to setup and integrate with existing institutional facilities
Ongoing costs	No ongoing subscription cost	Ongoing subscription fees per month/annum and/or per user
	No additional cost per additional user (assuming available server capacity)	Typically each additional user increases costs Increasing user data may result in additional subscription costs
	Periodic hardware maintenance cost (e.g. replacing servers, or increasing storage capacity) Time cost for regular system maintenance If using a private cloud-server, ongoing rental cost	No ongoing hardware/maintenance costs as encompassed in fee
	Time cost for training each additional new user Time cost for training and maintaining support staff for ELN infrastructure	Time cost for training each additional new user

Ongoing development	Reliant on open-source community to continue to maintain software (e.g. issuing patches for security vulnerabilities, or development of new features)	Software updates (security, new features) normally provided as part of subscription
	Users can directly suggest and implement new features	New feature development dependent upon company priorities
Software security	Public codebase allows interrogation from many parties, quickly identifying security issues	Private codebase limits scope for third-parties to identify security issues, reliant on the company to quickly implement fixes
Outboarding strategy	Free access to all notebook entries and files	If company fails, risk of data becoming trapped/inaccessible on company servers
	Either export feature incorporated or can create custom methods for extracting data	Reliant on software to provide appropriate export solutions – some may restrict file formats or the preservation of certain features or elements, leading to risk of data lock-in
Data storage	Complete control over location and accessibility of stored data	Product-dependent, some may offer compliance with specific regulations

Table 2: Comparison of ELN features for open-source and commercial solutions. Note: this is a generalised comparison – not all ELN products offer the same levels of functionality.

Electronic laboratory notebooks are not a solution to poor data management

ELNs do not resolve the challenge of systematically storing raw data and are just one part of a holistic data management strategy. Successfully implementing an ELN requires reflection on current practice to determine how a laboratory handles and stores information.[39,40] As with paper notebooks, policy, training and enforcement are required to ensure users record timely, useful and complete notebook entries. New users require training to understand the purpose and expectations of laboratory notebook use within a given organisation. A policy for how, when and by whom notebooks are monitored sets expectations. An offboarding procedure should be implemented to ensure outgoing user data is appropriately archived.

Procedures for linking raw data, laboratory notebook entries, analysed data, and publication data should be clearly defined and enforced. Data should be stored on centralised group or measurement- specific servers, or publicly-available repositories, with redundant backups. Persistent identifiers such as digital object identifiers (DOIs) can be used to help link resources, with ELN entries acting as centralised documents that connect to multiple files and data.[39] Some ELNs already generate unique identifiers for each notebook entry.[13] To help verify the integrity of externally stored data, ELNs can be used to record cryptographic hashes of data files.

Some ELNs feature application programming interfaces (APIs) that allow other software to directly read and write notebook entries.[13] For example, a user conducting a computer-controlled experiment could allow that equipment to directly record experimental or process details, potentially streamlining routine measurements. Laboratories with well-defined workflows, such as electron microscopy,[41] or genetic analysis,[30] may benefit from specialised ELNs that incorporate notebook entries within the data capture workflow, or which have been designed with equipment integration as a primary goal.

Successful ELNs recognise the needs of their users

Within an academic research environment both time and money are at a premium. Ideally, if adopted, ELNs for the academic research laboratory should be implemented at institutional level, harnessing existing ICT infrastructure, and with a sufficient commitment for ongoing support to encourage uptake.[42] While a handful of university-based surveys and studies of ELN implementation exist,[16,42,43] the reported level of success varies, suggesting careful user-engagement, product choice and ongoing support is key to successful deployment.

Critically, the implementation of an ELN should not introduce a net burden to the user. If users are unable or unwilling to use time-saving features, adopting an ELN may ultimately be a hindrance. For example, for a researcher who regularly draws chemical structures or writes equations into their notebook, most non-specialist ELNs will be less convenient than paper,[42] and the introduction of an ELN may be undesirable.[20] Lack of support for systems such as LaTeX can be a barrier to adoption in specific disciplines.[44] Recognising the needs of an often diverse range of researchers is essential before making decisions on the choice of software and approach.

Internal trialling of a small number of products, jointly agreed compromises or incorporating integrations with specialist solutions may be required, such as ELNs capable of handling chemical structures.[45] Simple infrastructure factors, such as unreliable WiFi,[7,16,27] insufficient benchspace to place a laptop computer,[16] or lack of access to up-to-date hardware and software in the laboratory can also adversely affect users.[42] The prevalent culture of private and personal academic laboratory notebooks should be recognised, with one ELN study

noting that researchers felt embarrassed when required to share their notebook with colleagues.[46]

User training should be recognised as an additional time burden, with researchers already expected to master a wide range of software packages.[47] Similarly, many laboratories already use laboratory information management systems (LIMS) for inventory handling, equipment booking, and procurement. Introducing a further standalone system can work against the time-benefits of digitisation and integration.[48] Implementing an ELN is an opportunity to reflect upon and consolidate existing practice, with many ELNs incorporating LIMS-style components. Similarly, if it is intended for the ELN to integrate with specific pieces of hardware, it cannot be assumed that software will work seamlessly, so system integration should be tested prior to deployment and the appropriate resources allocated to maintain the integration.

To date, there have been no published multi-year longitudinal studies of ELN implementation in academic environments. Hence, it is important to consider user issues that may appear over longer time scales. For example, determining how future researchers will be made aware of existing records if the original author (or their supervisor) has left the institution. Individuals should be identified who will facilitate access requests to existing records from new users. Investing in user training from the start should ensure that third parties are able to effectively locate these records in the future, i.e. that sufficient notation and metadata is being captured at the point of entry. Academic institutions may be able to take advantage of the existing in-house expertise of academic librarians, research data managers, compliance administrators, systems administrators and archivists, to help develop long-term workable policies. The general challenge of long-term digital data preservation and access exists within any large organisation (for example, email preservation), so leveraging existing good practice may aid deployment.

Box 1 describes our own experience of implementing an open-source ELN, illustrating some of the points discussed in the article. The choices made arose out of the specific set of circumstances appropriate to our research group, so this should not be considered a prescription for the best approach to implementing an ELN, but rather as a reflection on what we have learnt.

Box 1: Case study of implementing an open-source ELN

Since 2018, the authors have used eLabFTW as our ELN.[13] The criteria for identifying a suitable ELN focussed on accommodating scientists from multiple disciplines (e.g. cell biologists, synthetic chemists, engineers, etc) and the correspondingly diverse range of laboratory environments. Much of the research is applied, so reliable timestamping integration was considered essential for intellectual property protection. With a large number of active users at any given time, the system had to scale affordably, with persistent and ready long-term access to archived user data.

The large number of users meant the projected ongoing per-user subscription costs for commercial products were considerable. This potential cost combined with our other criteria, led us to select the open-source software eLabFTW after a successful 6-month trial prior to wider deployment. The server runs locally behind an institutional firewall on modest repurposed hardware. The trial period allowed a local ELN policy to be developed. No user is forced to use the ELN, existing lab members can switch to the system, and new members are offered the choice of electronic or paper notebooks. However, users may not operate both types concurrently to avoid data fragmentation. All new users go through an induction and training process. Compliance with the policy is checked periodically to ensure users are timestamping their entries (eLabFTW provides server and user level statistics to facilitate this).

Financial costs included purchasing an uninterruptable power supply (~£300), a network attached storage device (~£190, one component of a multi-site backup strategy) and a subscription to the third-party trusted timestamping authority (~£70 / year). Indirect costs included the time of a researcher to act as system administrator to setup and configure the server (taking advantage of existing expertise within the group), support from institutional ICT staff to support configuration and integration with existing institutional services, and the time cost of running regular user induction and training. Ongoing costs include administrating the server (~3 hours / month, performed internally by senior researchers within the research group) and training (~3 hours / month).

After approximately 2.5 years, ~60% of current lab members have active ELN accounts. Anecdotally, onboarding new students and staff as they arrive in the group appears to minimise the inconvenience of adopting an ELN, as they develop their new workflows concurrently with learning how to use the system, as well as adapting to the wider standards and protocols of the group surrounding research data management and reporting.

Challenges have included occasional software bugs, requiring a workaround until an update was available. A handful of network and power related issues have resulted in a total downtime of approximately 3 days over the course of 3 years. Ongoing user training has included reminders of the importance to timestamp entries. The generic nature of notebook entries has presented some challenges for computational scientists, where computational notebooks are limited to file attachments within the ELN, an inelegant solution for those used to working with version control systems such as Git. Understanding how new users use the ELN is an ongoing challenge – we have found using a dedicated communications channel in Microsoft Teams has aided this process. Long term the ELN is reliant on having suitably qualified researchers able to administer the system (one of the costs of an open-source solution).

The advantages have included the ability to readily share lab notebook entries between users, for example protocol sharing within the same project and providing new students and researchers with direct access to the work of former group members. This has included during the COVID-19 pandemic, where users are able to access the ELN from home.

Students are able to share notebook entries directly and remotely with senior researchers, meaning they can efficiently receive guidance, again useful in a period where lab access has been restricted. We have observed that since the ELN has been introduced, the user base has grown spontaneously with awareness of the system.

END BOX

ELNs are an opportunity to consider the broader philosophy of academic laboratory research

With care, ELNs can be used to support information capture, making it more consistent, accessible and usable to both current and future generations of scientists. Implementing an ELN provides an opportunity to consider how digitised notebook entries can help address some of the broader challenges of academic research. For example, some ELNs cater to aspects of open science,[49,50] allowing them to be configured to share data outside of organisations, supporting initiatives such as the FAIR Data Principles (a proposal that scientific data should be Findable, Accessible, Interoperable, and Reusable).[51] Similarly, the integration of ELNs with institutional repositories may offer new opportunities to tackle the challenge of research data management, such as efforts by the University of Edinburgh to allow users to directly deposit data into an institutional repository via their ELN interface.[52]

Adopting an ELN abstracts the user from the underlying notebook storage technology, allowing a wide-range of other approaches to be implemented, different to the one illustrated in **Figure 3**. Rather than using a private database of notebook entries, an ELN could implement a publicly distributed decentralised record store, using technologies such as blockchain and peer-to-peer networking,[53] to aid accountability or reduce the reliance on only one repository for long-term data storage. For example, the research project bloxberg is a blockchain operated by an academic consortium,[54] which has been independently integrated into the ELN eLabFTW.[13] Cryptographic hashes of notebook entries can be timestamped and recorded on a publicly accessible ledger, distributed across the consortium network, thus removing the need for a single trusted third-party timestamping authority. In effect, verification of the integrity of notebook entries (or other scientific data) is distributed amongst the consortium members.

As ELNs provide the primary interface to research data, there is the opportunity to consider how they can integrate with computational approaches that aim to automatically derive insights into the data. This could include the integration of an ELN with computational semantic technologies,[20] which allow the meaning of human language to be automatically inferred. This would allow research metadata to be automatically derived, aiding search efforts; or create automated insights by linking relevant data together.[48] Alternatively, an ELN might be integrated with existing community-defined ontologies and databases for specific disciplines. In this scenario the ELN would act both as a form of input validation, ensuring data is captured according to community standards, and as a mechanism for facilitating access to underlying data to facilitate large scale meta-analyses.

In summary, for researchers and institutions considering implementing an ELN, a nuanced understanding of laboratory culture is needed to facilitate a respectful and ultimately user-supported deployment. The barriers to entry must be carefully managed as these have the potential to create technological divides within the academic community, diluting the benefits of ELNs. Successful ELN implementation should be seen as an ongoing commitment to ensure the needs of different users continue to be met. Given that the current median lifetime of ELN software packages is 7 years, it is of utmost importance to have sufficient ongoing institutional support to maximise the value to researchers and mitigate the risks, enabling continuity when software changes are required. With careful consideration, successful implementation of an ELN presents a pathway to greater knowledge development and transfer within academic research.

Supporting Information

The associated Supplementary Methods for this manuscript contains a detailed description of the ELN product survey methodology.

Data Availability

The raw data, containing the product names, manufacturers, estimated release and defunct dates and corresponding web references to support these estimates, along with the R Notebook used to generate the figures, and an interactive version on Figure 1, is publicly available from a data repository (Zenodo).[17] The analysis used various R software packages.[55–59]

Acknowledgments

Author contributions

S.G.H. and A.N.V. wrote the manuscript. M.M.S. supervised the work and edited the manuscript.

Competing interests

S.G.H. and A.N.V. are the system administrators for the Stevens Group ELN, which uses the open-source software package eLabFTW. S.G.H. has provided feedback via public forums to the developer of eLabFTW.

References

1. White, K. *Publications Output: U.S. Trends and International Comparisons | NSF - National Science Foundation*. https://ncses.nsf.gov/pubs/nsb20206/ (2019).

2. Waldo, W. H. & Barnett, E. H. An Electronic Computer as a Research Assistant. *Ind. Eng. Chem.* **50**, 1641–1643 (1958).

3. Borman, S. A. Scientific Software. *Anal. Chem.* **57**, 983A-994A (1985).

4. Gilbert, W. A. RS/1: an electronic laboratory notebook. *BioScience* **35**, 588–590 (1985).

5. Schapira, M., The Open Lab Notebook Consortium & Harding, R. J. Open laboratory notebooks: good for science, good for society, good for scientists. *F1000Research* **8**, 87 (2019).

6. Williamson, A. E. *et al.* Open Source Drug Discovery: Highly Potent Antimalarial Compounds Derived from the Tres Cantos Arylpyrroles. *ACS Cent. Sci.* **2**, 687–701 (2016).

7. Adams, R. *et al.* Digital Notebooks - productivity tools for researchers - event materials. (2018) doi:10.5281/zenodo.1247390.

8. Downie, A. Electronic Lab Notebooks - for prospective users. *The Gurdon Institute* https://www.gurdon.cam.ac.uk/institute-life/computing/elnguidance (2020).

9. Douglas, S. ELN vendor. *LIMSWiki* https://www.limswiki.org/index.php/ELN_vendor (2021).

10. Harvard Medical School. Electronic Lab Notebooks. *Harvard Medical School* https://datamanagement.hms.harvard.edu/electronic-lab-notebooks (2021). **At the time of writing, this is a regularly maintained and comprehensive comparison grid of more than 30 ELN products, presented from an academic perspective.**

11. Ritt, S. & Paul Scherrer Institut. ELOG Linux Demo. https://elog.psi.ch/elogs/Linux%20Demo/ (2021).

12. Textco BioSoftware, Inc. Gene Inspector (GI) | Electronic Laboratory Notebook | Sequence Analysis Software | Textco BioSoftware. http://www.textco.com/gene-inspector.php.

13. CARPi, N., Minges, A. & Piel, M. eLabFTW: An open source laboratory notebook for research labs. *J. Open Source Softw.* **2**, 146 (2017).

14. Bowman, S., Pfeiffer, N. & Pfeiffer, N. *Example project for demonstration: Hydroxymethylfurfural oxidation over supported metal catalysts.* https://osf.io/xj7a3/ (2019).

15. Rubacha, M., Rattan, A. K. & Hosselet, S. C. A Review of Electronic Laboratory Notebooks Available in the Market Today. *JALA J. Assoc. Lab. Autom.* **16**, 90–98 (2011).

16. Kanza, S. *et al.* Electronic lab notebooks: can they replace paper? *J. Cheminformatics* **9**, 31 (2017). **This is a comprehensive and balanced study on the implementation of ELNs in an academic setting, exploring barriers to entry and adoption.**

17. Higgins, S. G., Nogiwa-Valdez, A. A. & Stevens, M. M. Research data supporting 'Considerations for Implementing Electronic Laboratory Notebooks in an Academic Research Environment'. *Zenodo* https://doi.org/10.5281/zenodo.5012729 (2021).

18. Dessy, R. Electronic Lab Notebooks. *Anal. Chem.* **67**, 428A-433A (1995).

19. Butler, D. A new leaf. *Nature* **436**, 20–21 (2005).

20. Kanza, S., Gibbins, N. & Frey, J. G. Too many tags spoil the metadata: investigating the knowledge management of scientific research with semantic web technologies. *J. Cheminformatics* **11**, 23 (2019).

21. Fowler, J. *et al.* Experience with the virtual notebook system: abstraction in hypertext. in *Proceedings of the 1994 ACM conference on Computer supported cooperative work* 133–143 (Association for Computing Machinery, 1994). doi:10.1145/192844.192892. **This is a report on experiences with a relatively early ELN system back in 1994, that (amongst the technical description) highlights**

key lessons, such as the challenge of trying to apply a uniform system to heterogeneous groups of academics.

22. British Standards Institution. *BS EN ISO 15189:2012: Medical laboratories. Requirements for quality and competence*. (BSI, 2012).

23. British Standards Institution. *BS EN ISO/IEC 17025:2017: General requirements for the competence of testing and calibration laboratories*. (BSI, 2017).

24. U.S. Food and Drug Administration. *Code of Federal Regulations Title 21 Part 11. Code of Federal Regulations* vol. 21.

25. Gorry, G. A. A Virtual Notebook for Biomedical Work Groups. *Bull Med Libr Assoc* 12 (1988).

26. Hansen, J. D. & Reich, J. Democratizing education? Examining access and usage patterns in massive open online courses. *Science* **350**, 1245–1248 (2015).

27. Bromfield Lee, D. Implementation and Student Perceptions on Google Docs as an Electronic Laboratory Notebook in Organic Chemistry. *J. Chem. Educ.* **95**, 1102–1111 (2018).

28. Tremouilhac, P. *et al.* Chemotion ELN: an Open Source electronic lab notebook for chemists in academia. *J. Cheminformatics* **9**, 54 (2017).

29. Shipman, F. M., Chaney, R. J. & Gorry, G. A. Distributed hypertext for collaborative research: the virtual notebook system. in *Proceedings of the second annual ACM conference on Hypertext* 129–135 (Association for Computing Machinery, 1989). doi:10.1145/74224.74235.

30. Bauch, A. *et al.* openBIS: a flexible framework for managing and analyzing complex data in biology research. *BMC Bioinformatics* **12**, 468 (2011).

31. Guerrero, S. *et al.* Analysis and Implementation of an Electronic Laboratory Notebook in a Biomedical Research Institute. *PLOS ONE* **11**, e0160428 (2016).

32. European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Off. J. Eur. Union* **59**, 2016/679 (2016).

33. Imperial College London. Storing sensitive and personal data. *Imperial College London* http://www.imperial.ac.uk/research-and-innovation/support-for-staff/scholarly-communication/research-data-management/data-storage-and-security/storing-sensitive-and-personal-data/ (2021).

34. Gotthardt, M. A flexible Electronic Lab Notebook (ELN) with Microsoft OneNote. *The LabOMator* http://lab-o-mator.blogspot.com/2015/06/a-flexible-electronic-lab-notebook-eln.html.

35. Adams, C. & Pinkas, D. Internet X.509 Public Key Infrastructure Time-Stamp Protocol (TSP). https://tools.ietf.org/html/rfc3161.

36. American National Standards Institute. *ANSI X9.95-2016 - Trusted Time Stamp Management and Security*. (Accredited Standards Committee X9, Incorporated, 2016).

37. University of Oxford. More About Electronic Lab Notebooks. *University of Oxford* https://help.it.ox.ac.uk/research/eln/moreaboutelns (2018).

38. Foster, E. D. & Deardorff, A. Open Science Framework (OSF). *J. Med. Libr. Assoc.* **105**, 203–206 (2017).

39. Gerlach, B., Untucht, C. & Stefan, A. Electronic Lab Notebooks and Experimental Design Assistants. in *Good Research Practice in Non-Clinical Pharmacology and Biomedicine* (eds. Bespalov, A., Michel, M. C. & Steckler, T.) 257–275 (Springer International Publishing, 2020). doi:10.1007/164_2019_287.

40. Brown, C. Electronic Lab Notebooks: Solutions for Paperless Research. *Research infrastructure and data* https://researchdata.jiscinvolve.org/wp/2017/01/18/eln/ (2017).

41. Rees, I., Langley, E., Chiu, W. & Ludtke, S. J. EMEN2: An Object Oriented Database and Electronic Lab Notebook. *Microsc. Microanal.* **19**, 1–10 (2013).

42. Vaas, L. A. I. *et al.* Electronic laboratory notebooks in a public–private partnership. *PeerJ Comput. Sci.* **2**, e83 (2016).

43. Dirnagl, U. & Przesdzing, I. A pocket guide to electronic laboratory notebooks in the academic life sciences. *F1000Research* **5**, 2 (2016).

44. Magid, A. Is It Worth It? Implementation of Electronic Lab Notebook Software Among the STEM Community at an American University in the UAE. in *2018 ASEE Annual Conference & Exposition* (American Society for Engineering Education, 2018). doi:10.18260/1-2--30734.

45. Kotov, S., Tremouilhac, P., Jung, N. & Bräse, S. Chemotion-ELN part 2: adaption of an embedded Ketcher editor to advanced research applications. *J. Cheminformatics* **10**, 38 (2018).

46. Neylon, C. Open Notebook Science: Perspectives from a newbie. *Nat. Preced.* 1–1 (2007) doi:10.1038/npre.2007.1130.1.

47. Lysakowski, R. The CENB Advantage. *Bio/Technology* **13**, 347–348 (1995).

48. Talbott, T., Peterson, M., Schwidder, J. & Myers, J. D. Adapting the electronic laboratory notebook for the semantic era. in *Proceedings of the 2005 International Symposium on Collaborative Technologies and Systems, 2005.* 136–143 (2005). doi:10.1109/ISCST.2005.1553305.

49. Switters, J. & Osimo, D. *Electronic Laboratory Notebooks (ELNs) as key enablers of open science: open science monitor case study.* (European Union, 2019).

50. FOSTER consortium. *What is Open Science?* https://zenodo.org/record/2629946#.YC-6i2OnxTY (2018) doi:10.5281/zenodo.2629946.

51. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 1–9 (2016).

52. Macdonald, S. & Macneil, R. Service Integration to Enhance Research Data Management: RSpace Electronic Laboratory Notebook Case Study. *Int. J. Digit. Curation* **10**, 163–172 (2015). **This article is a case study illustrating how ELNs can be integrated directly with a university data repository, allowing users to deposit data through the ELN for compliance with an RDM policy.**

53. Bartling, S. Blockchain for Open Science - the living document. *Blockchain for Science* https://www.blockchainforscience.com/2017/02/23/blockchain-for-open-science-the-living-document/ (2017).

54. Kleinfercher, F., Vengadasalam, S. & Lawton, J. *bloxberg, The Trusted Research Infrastructure, Whitepaper 1.1.* https://bloxberg.org/.

55. R Core Team. *R: A Language and Environment for Statistical Computing.* (R Foundation for Statistical Computing, 2020).

56. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).

57. Kirill Müller. *here: A Simpler Way to Find Your Files.* (2020).

58. Sievert, C. *Interactive web-based data visualization with R, plotly, and shiny.* (Chapman and Hall/CRC, 2020).

59. Vaidyanathan, R. *et al. htmlwidgets: HTML Widgets for R.* (2020).
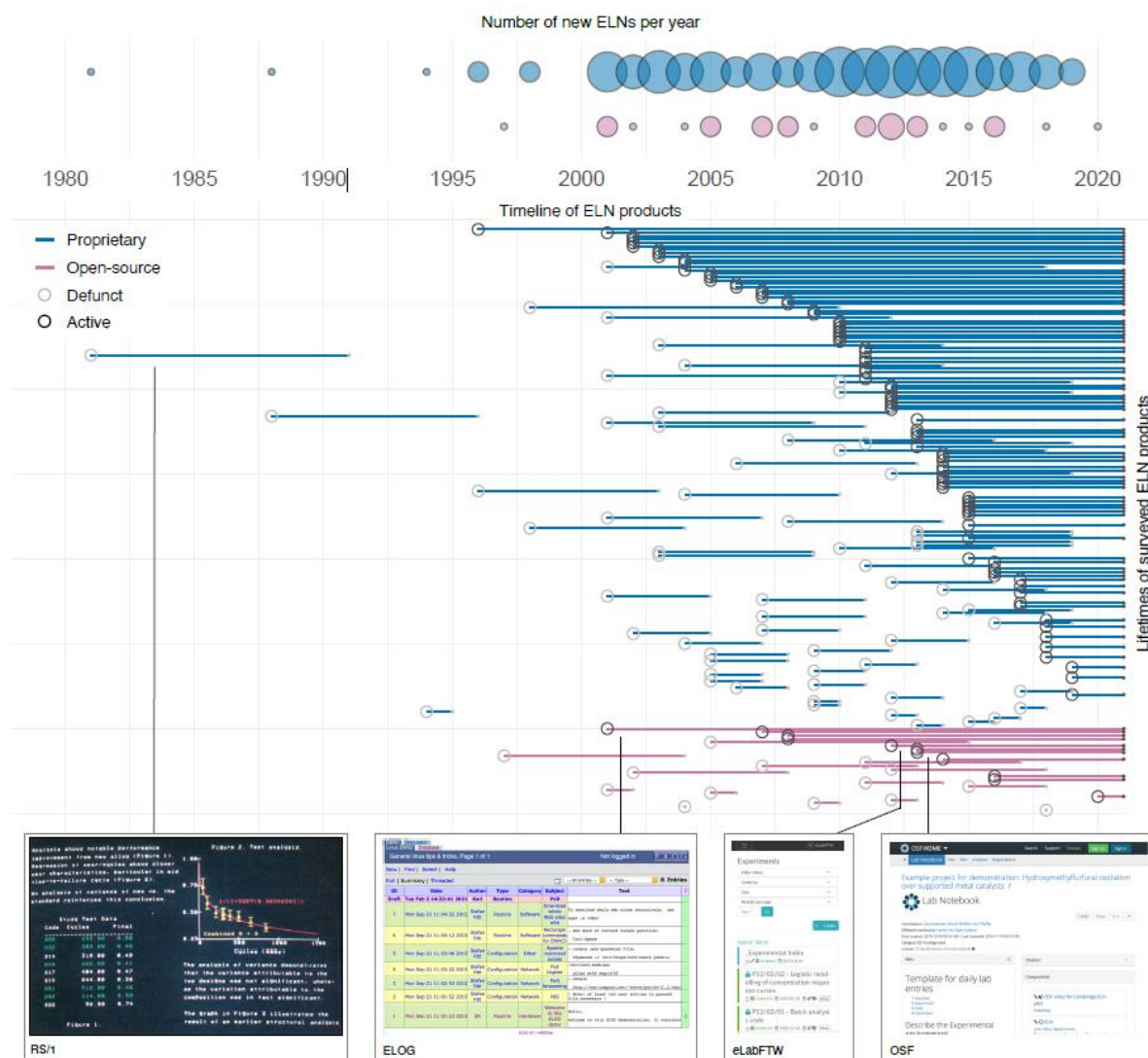
Figure Legends

Figure 1: Timeline of 172 ELN software packages, documenting known or estimated release and cessation year. The area of the circles shown at the top of the figure is proportional to the number of new ELNs launched in a given year. Data is segregated into proprietary (blue) and open-source (pink) codebases, and sorted within these categories according to the number of years the software has been active. Insets ar screenshots of a selection of ELNs, showing the progression of software development with time, including: RS/1 (Bolt Beranek and Newman Inc), reprinted (adapted) with permission from Borman et al.[3] Copyright (1985) American Chemical Society. ELOG (Paul Scherrer Institute), captured from online source and reproduced with permission.[11] eLabFTW (Deltablot),[13] captured from authors' own server and reproduced with permission. OSF (Centre for Open Science), captured from online source,[14] and reproduced under the terms of a CC-BY 4.0 license. Copyright (2019) The Authors. See Supplementary Method 1 for a description of the survey methodology and limitations. This plot incorporates data from primary and secondary sources.[9,15,16] An interactive version of this figure, along with any updates, is available from a data repository (Zenodo) supporting this manuscript.[17]
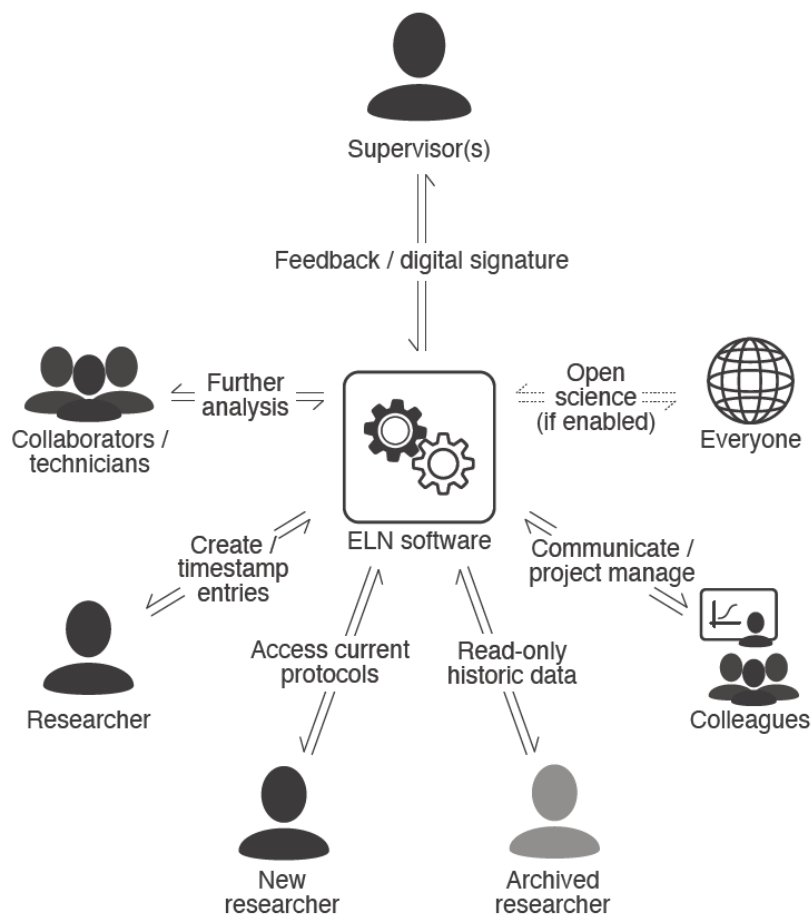
Figure 2: Examples of possible information flows between different users, as mediated by an ELN. Researchers can create and secure entries, share information with colleagues and collaborators and access the records of former team members. The degree of sharing is dictated by local policy, software features and configuration.
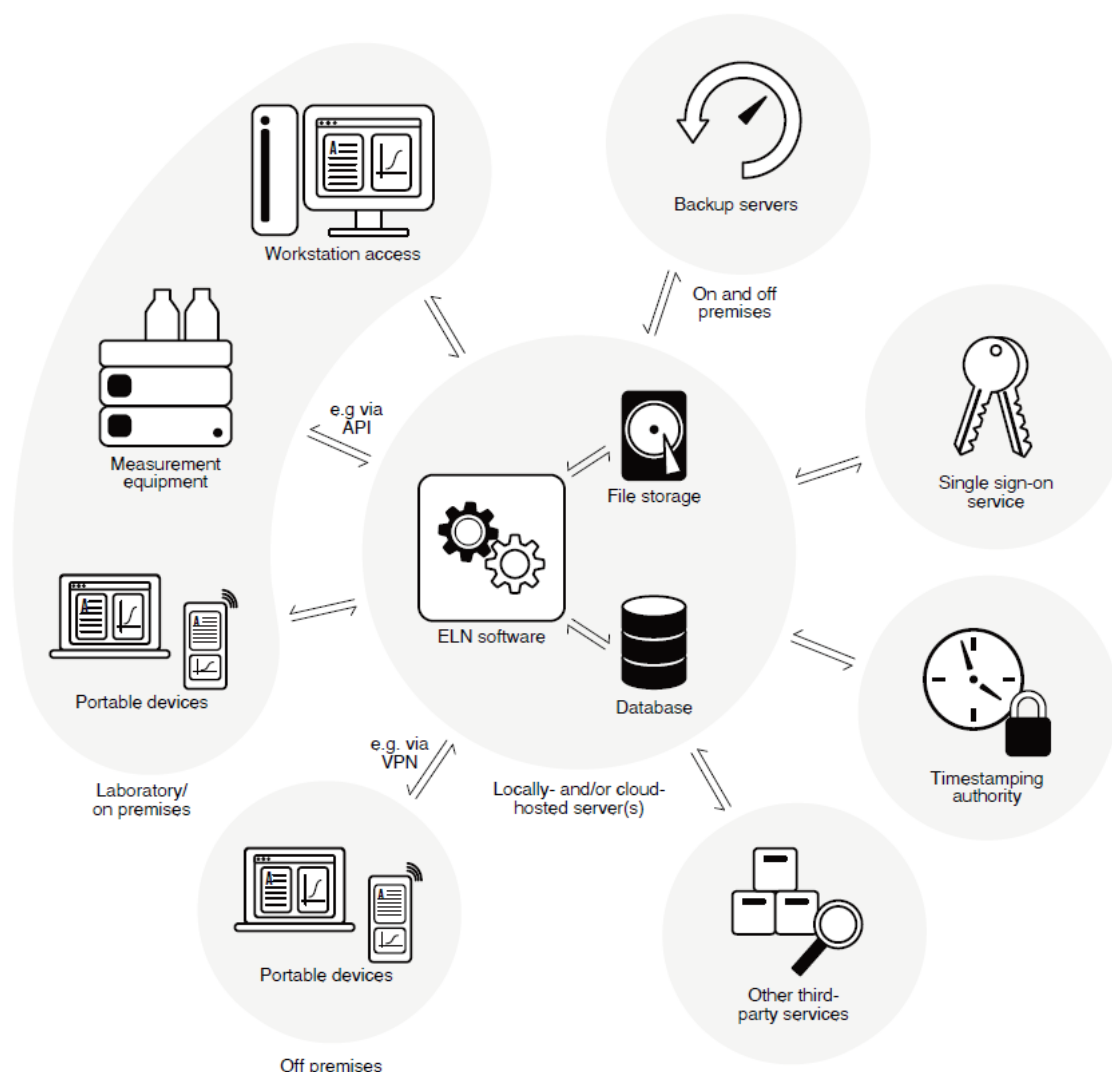
Figure 3: Simplified illustration of a commonly used ELN architecture. Users interact with the ELN software via workstations and/or mobile devices, with off-premises access possible via technologies such as virtual private networks or by exposing the ELN to the wider internet (the latter introducing additional security concerns). Experimental hardware may interact directly with the ELN server, for example via an application programming interface (API). ELN software, database and file storage may coexist on the same physical server or operate on separate hardware. The server may be run as a local installation at an institution, or as part of a cloud-hosted product from the software vendor. Additional requirements include backup servers, which should be geographically separate from the ELN (and should consist of multiple layers of redundancy). Third-party services include access to single sign-on servers (to facilitate a better user experience), trusted third-party timestamping authorities (for independently digitally signing notebook entries), or other laboratory systems (such as laboratory inventory management systems, etc).