Digital Lab Notebooks - accessible tools in the practice of science research

Daniel Chin and Richard Badge, Life Sciences

Abstract

Lab Notebooks are records of research that are required for laboratory scientists to create, repeat and protect their research. Computer-based Digital Lab Notebooks, or DLNs, are growing in usage with cheaper and faster Information and Communication Technologies (ICT), to the point where tablet computers can speedily download protocols and upload results securely. Advantages of DLNs include speeding up editing processes, creation of definitive records with time-stamps, and the potential to bring a greater diversity of researchers into STEAM with more accessible methods for making records. Implementing such tools necessarily involves teaching individuals how to use them, as well as sufficient infrastructure to support the technology. We have reviewed the landscape for DLN tools that are free to use, discussed their usability in an academic environment, and considered how they could be improved.

Introduction (lab notebooks and digital lab notebooks)

Scientific research is an evidence-based practice. Research requires utilising procedures in experiments, which involve creating and following a set of instructions to produce an observed outcome. In order to empirically support the validity of one's observations borne from research, records of experiments, while they are being performed, are essential. The core content in a Lab Notebook is documenting how you performed an experiment (methods), the product of said experiment (results) and your interpretation of what the data means, to lead on to further experiments (discussion/further work).

Not only is it good practice to record information directly from results, but detailed records also allow other researchers to repeat your experiments enhancing reproducibility, in the absence of direct observation of your results. In science today, this is also important for establishing the timing of discoveries, primarily because research is increasingly a commodity, funded by businesses or governments. This means that such records require legal and physical protection to support claims of intellectual property. This is the basis for the practice of keeping a record of research in the form of a Lab Notebook. As recording tools, notes of research have been kept in permanently bound paper books for decades. The earliest and most notable example of this practice date was the research notebook of Alexander Graham Bell (Bell *et al.,* 1875). His extensive record keeping of how he invented the telephone showed firm evidence of originality, thus allowing the invention to be patented.

Technological advances have enabled information to be recorded and stored digitally. This has given rise to the Digital Lab Notebook (DLN). These are alternatives to traditional pen-and-paper Lab Notebooks that use a computer program within an electronic device to function as a recording device. The main advantages of using digital technology is being able to securely store and communicate your results effectively. This advantage comes in the form of being able to type lab notes, removing issues relating to legibility of hand-written notes, as well as being able to easily edit mistakes. Furthermore, being able to copy sections for re-use (such as a standard protocol) allows greater efficiency, particularly if one wants to repeat, but only change small parts of a method. Problems with portability and usability within lab settings are gradually

being reduced. For example, during my final year research project, I was able to bring a protected iPad into a genetics laboratory, to refer to protocols during PCR reaction setup, for example, and to record the results of electrophoresis experiments as images. Of course, the utility of freely editing a DLN could also be a potential problem when ensuring the validity of record keeping – preserving audit trails of note editing is integral to DLNs. Furthermore, as DLNs exist only as digital data, security of data storage and backup is paramount, especially with increasing use of de-localised storage, such as cloud servers. These technical issues are compounded with the requirement for basic computer literacy to enable use of DLNs in the most effective manner. This may be part of the problem preventing DLNs being widely implemented, as previous practices still prevail – DLNs are a fairly new technology, but with much potential to grow.

While they are an essential part of research in industry, we have yet to see them being used or taught as core curriculum in the University of Leicester Biosciences courses, but there is scope to change this. A central question is "What are the challenges with implementing DLNs in academia?" We aim to obtain an overview of the landscape of DLNs currently available, explore the features of these DLNs and use this as a basis to discuss how DLNs could be improved, to ensure they are fit for purpose, in a practical lab setting.

Methods

Research of DLNs was done on an observational case-by-base basis. Various DLN softwares were tested for usability within a lab setting. Other technology was also researched for usability for students with special educational needs. This involved consulting with the AccessAbility Service for Disability Student allowance (DSA) support equipment, and an AccessAbility tutor with previous experience as an academic in the Biological Sciences. Independent research on voice-based solutions, using Amazon Alexa were also conducted.

Results

Collation of properties investigated DLN platforms

The case-by-case analysis results were collated against a set of aspects considered important for implementation in an academic environment (Table 1, columns a-g). These aspects arose from the experience of using the Mediawiki service at Leicester, for digital note-taking in a third year Undergraduate Research Project. As a result the Mediawiki was used as a benchmark for comparison.

Table 1: a comparison of all DLNs investigated in this research project. Each DLN is compared based on a) how much it costs to use, b) the operating system, c) if it has any pre-made templates, d) what media one can put into the DLN, e) how secure the service is, f) any notable extra features and g)notable problems with usage.

Software	**a) Cost Model**	**b) OS**	**c) Templates**	**d) Media input**				**e) Security**	**f) Extra features**	**g) Problems**
				Text	**Photo**	**Audio**	**Video**			
MediaWiki (Mediawiki.org, 2019)	Free open source, Server costs (free for academics)	HTML (i.e. any web browser)	None out the box – requires curation	Yes	Yes	Yes	Yes	Backed up and protected on-site by UoL	HTML-based, universal. Virtual web server protection	Requires knowledge of HTML to use most effectively
One Note (Microsoft, 2018)	Subscription, freemium (for University staff and students)	Windows, iOS, android	Yes (none specialised for lab work)	Yes	Yes	No	No	Autosave backups, off-site	Useful on a variety of platforms, ability to draw freehand	Lacks time stamps – ineffective recording, limited portability
Evernote (Evernote Corporation, 2018)	Freemium, Subscription	Windows, iOS, android	Yes (non-specialised)	Yes	Yes	Yes	No	Backed up to servers (off-site)	Webpage screenshotting, document scanning, tags, syncing across multiple devices, searching (pdfs, document & handwriting)	Lack of security (3rd party assured)
HiveBench (Elsevier, 2019)	Freemium	iOS	Yes (PCR, reagents)	Yes	Yes	No	No	Yes (off-site)	Used in a variety of industry and academic settings	Expensive, iOS exclusive
Findings (SAS, 2018)	Freemium	iOS	Yes (limited)	Yes	Yes	No	No	Yes (dropbox)	IOS exclusive	IOS exclusive
MySQL (Oracle Corporation, 2019)	Subscription only	HTML	Yes	Yes	Yes	No	No	Yes	Open Source, malleable, secure	Significant barrier for entry (learning)
Blackboard (Blackboard Inc., 2014)	Subscription Only (via University account)	HTML	No	Yes	Yes	No (by default)	No (by default)	Backups (provided saved)	Part of Blackboard VLE service	Future portability of journal entries

Benchmark: MediaWiki

During my DIP research, I used the *Badgelab MediaWiki* DLN [Available at: https://wiki.lamp.le.ac.uk/badgelab/index.php/Main_Page BUT IT account, permissions required] to record my research notes and results for my final year research project (See Figure 1). It is like a private Wikipedia website for personal lab notes. It can host many file formats on the site, including text, images, audio and video. Any scientific research project requires methods, results and interpretation to be conveyed through text-based documentation. My project benefitted from the other formats, as I could record images of agarose electrophoresis gels, to use as records of results. This DLN was hosted on the University of Leicester's server within the UK. Immediately, this means that costs for running it on an online server are absorbed by the University, making it free for users such as myself. Security costs are also minimised, as the University's IT Service ensures that all online materials and systems are protected.

HTML is the language that MediaWiki uses to make pages. The main advantage of using this open-source language is its portability. Any device with an internet connection and an internet browser can understand HTML and produce a readable output. ICT has got to a point where I could access my DLN via a tablet computer. When protected, I could perform my PCR and electrophoresis experiments accurately by referencing my notes in the lab. Not only could I immediately record my results, but I could also make amendments to my protocols if I made an error, with all changes backed up to a central server. Furthermore, transporting files is easy, as it is possible to copy and paste the source code between MediaWikis. Transporting source code is crucial for keeping research notes, as being able to access your research with future proofing is essential.

There was a steep learning curve in attempting to use it, as I have not been trained in the language. Fortunately, there were plenty of resources available in the form of University guideline pdfs and some easy short-cut buttons to implement formatting (such as fonts and hyperlinks). Other people also use the wiki to host their notes, meaning that I could access their MediaWiki code to copy tables and protocols. This significantly aids in collaboration between colleagues, directly through shared notes, but also indirectly by using templates that were created and refined over many years of usage.

Finally, time stamped backups are the most crucial part of a DLN which ensure it is an accurate record of Lab events. Every time you save a page, a version of the page is stored the wiki, with a record of the time and date it was saved. With multiple iterative records, it is easy to compare changes made to pages, allowing one to revert to a previous version of the DLN if there was accidental deletion / corruption of information. Theoretically, the timestamp can only be changed by the external administrator who isn't part of the lab, which reduces potential for erroneous time recordings. This system ultimately protects data from editing, as well as ensuring a paper trail for any edits to DLNs.

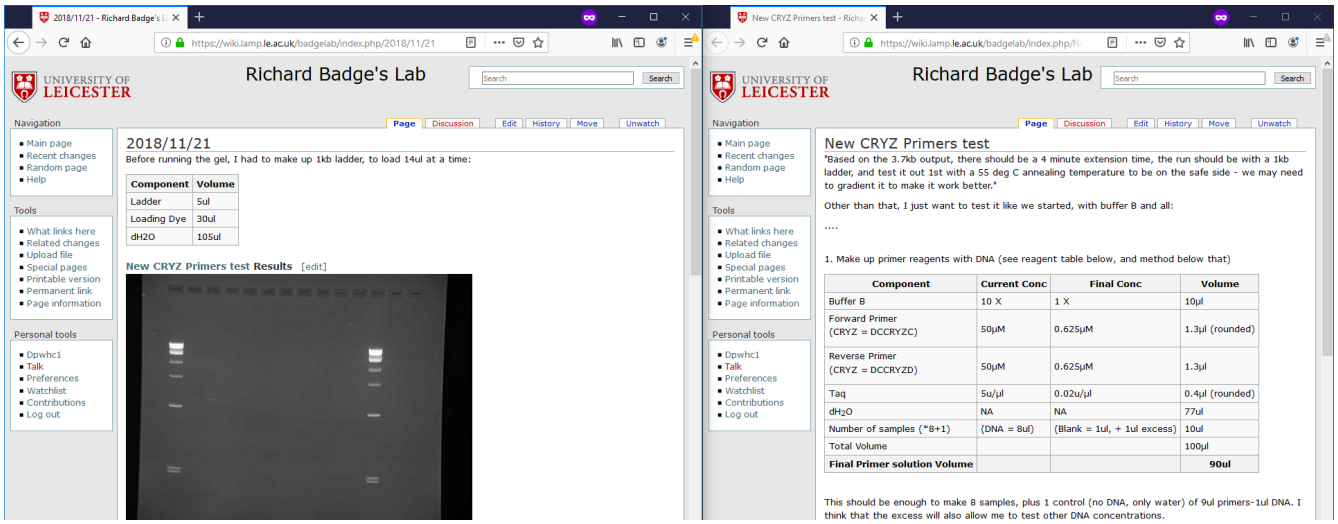

Figure 1: Badgelab MediaWiki screen shots. Example entries from the notebook used in final year project of Chin D.

Commercial DLN's

Other DLNs were compared to *Badgelab MediaWiki* for usability, as summarise in Table 1. *Microsoft's OneNote*, which was designed to be used as a digital notebook, has an easier to understand user interface, being able to write text and insert images, just like in Microsoft Word. It also had the advantage of both local and online storage of entries, which allows access to entries both online and offline. OneNote also includes some built-in templates, but none were specialised for bioscience laboratory usage. Unfortunately, the major problem was that storage of information is based on Back end protection. Online data is backed up offsite from the application supplier, in a country outside the UK. Relying on an external source for data protection may potentially cause security issues, especially where there are concerns over ownership rights. Furthermore, the multiple versions of copies both online and offline mean that reintegrating them together affects the timestamps. This means that there is a significant problem with the recording keeping aspect of Onenote. Backups are also restricted among the DLN's tested. *Findings* and *Evernote*, while they have a small amount of free backup storage (megabytes), both requires a subscription to acquire storage to be effectively used as a DLN (gigabtyes). Blackboard is the closest analogue to *Badgelab,* that is already widely implemented in Bioscience. Within the VLE, there are HTML-based pages one can create within certain tools (Blogs, for example). Most crucially, timestamps of saved entries are separate from the HTML code that protects any edits, making them sufficient permanent records.

Discussion

To summarise, most commercial DLN applications had significantly lower barriers of entry for usability than *BadgeLab,* which is ideal for first time users. However, none of them had the same functionality nor a comprehensive security solution suitable for an academic institution (on-site storage). A centralised model, like the one used in *Badgelab MediaWiki* and *Blackboard* can resolve the issues of timestamping because there is a single accessible copy linked to multiple versions of the document. However, both *MediaWiki* and *Badgelab* require a constant internet connection to use and cannot be stored offline. There is currently no standard DLN tool used across the University of Leicester. Fortunately, the Bioscience Department is planning on using the *Blackboard posting system*, and testing out journal-styled entry logs in final year projects (Suter-Giogini, 2019, personal communication, 16 May). At the core of any lab notebook, one needs

to be able to record research process – from my research, HTML-based approaches are more useful in teaching these essential skills.

Limited accessibility was a consistent pattern among all the commercial DLNs tested. Visual aspects, such as background colour, adjustable text size, are not built in and require the usage of external tools that are built into iPad devices. Visual design for visual executive dysfunction problems like Autism Spectrum Disorder (National Autistic Society, 2016) are present in the consumer-based designs. The HTML-based DLNs do not have this built in and require templates to be made with these considerations. Beyond visual limitations, those who are not computer literate or have significant sight impairments lack the means to use any of these tools. This could be ameliorated with the use of voice-controlled tools that can time experiments, calculate molarity and bring up relevant information such as restriction-enzyme digestion data (Biolabs, 2019). Unfortunately, we are not at the stage where we can dictate notes that directly convert voice to text, but natural language processing could make this possible.

I believe that DLNs can not only ease recording of research, but crucially aid in presenting them, allowing for greater understanding of information, leading to better outcomes for both scientist and society.

References

Bell, A., Watson, T., Hubbard, G. and Monroe, L. (1875). *Notebook by Alexander Graham Bell, from 1875 to 1876*. [online] The Library of Congress. Available at: https://www.loc.gov/item/magbell.25300201/ [Accessed 15 Jul. 2019].

Blackboard Inc. (2014). *Blackboard Learn | Learning Management System*. [online] Blackboard.com. Available at: https://www.blackboard.com/blackboard-learn/index.html [Accessed 26 Jul. 2019].

Biolabs, N. (2019). *myNEB™ | NEB*. [online] International.neb.com. Available at: https://international.neb.com/myneb/myneb [Accessed 15 Jul. 2019].

Elsevier (2019). *Hivebench*. [online] Hivebench. Available at: https://www.hivebench.com/ [Accessed 26 Jul. 2019].

Evernote Corporation (2018). *Best Note Taking App - Organize Your Notes with Evernote*. [online] Evernote. Available at: https://evernote.com/ [Accessed 26 Jul. 2019].

National Autistic Society (2016). *Autism-friendly websites - National Autistic Society*. [online] Autism.org.uk. Available at: https://www.autism.org.uk/professionals/others/website-design.aspx [Accessed 15 Jul. 2019].

Oracle Corporation (2019). *MySQL*. [online] Mysql.com. Available at: https://www.mysql.com/ [Accessed 26 Jul. 2019].

SAS, F. (2018). *Findings - All Your Research, In One App*. [online] Findingsapp.com. Available at: https://findingsapp.com/ [Accessed 26 Jul. 2019].

Wikimedia Foundation (2019). *MediaWiki*. [online] Mediawiki.org. Available at: https://www.mediawiki.org/wiki/MediaWiki [Accessed 26 Jul. 2019].