

Channelling information flows: a young researcher's approach to knowledge management

Based on a paper presented at the 37th UKSG Annual Conference, Harrogate, April 2014

As the volume of scientific literature continues to grow at an impressive pace, so too does the diversity of tools and approaches that are being proposed to track this knowledge. In the midst of such a vast ecosystem of journals and software, identifying meaningful innovation, as well as continuing to innovate, is essential to the emergence of powerful, flexible and easy-to-use 'channels' which researchers need to navigate the current scientific literature. In this article, the author uses the example of his own information consumption habits to suggest what he sees as the key components of such channels.

Discussion

Never has information been so abundant as it is today. A deluge of data is generated and stored in computers every single day, such that 90% of the world's data are less than two years old¹. Although this exponential growth of information is mostly driven by user-generated content on the internet, scientific literature has also been growing in a fast-paced manner, with nearly one million new papers being indexed on MEDLINE this year², and about three million connections being made every day on PubMed³. This is the context in which today's academics have to identify, reference and use the literature most appropriate to their needs.

Unsurprisingly, digital tools play an essential role in the everyday routine of academics and are an inseparable component of scholarly activity. Such tools hold the promise of improving the discoverability and accessibility of relevant literature, ultimately helping scientists to spot important discoveries, adapt their experimental design, solidify their interpretations in the light of similar work, and as a result ensure the production of high quality research. My impression, however, is that academics often find themselves confused by (or sometimes ignorant of) the plethora of tools, databases, public repositories and catalogues that populate the digital knowledge ecosystem. In addition, students and tech-savvy academics show a certain appetite for so-called 'productivity apps' that are, in most cases, targeted at a general audience, but which happen to somewhat suit academics' needs. This raises an important question: should we assume that the abundance of software and tools available to academics is sufficient for them to pave their way towards a fully digital and efficient scholarly activity?

Providing an honest and educated answer to this question requires an examination of three specific problems. Firstly, what are the challenges and needs of academics with respect to digital information consumption? Secondly, what tools are available that might match these needs? And, thirdly, how can we build efficient academic workflows? Interestingly, these questions might be addressed from two distinct standpoints: that of librarians and publishers on the one hand, and that of academics themselves on the other. I will attempt to bring a brief perspective on these problems on the basis of my own experience.



GUILHEM
CHALANCON

Gates Cambridge
Scholar, PhD Student
University of
Cambridge / MRC
Laboratory of
Molecular Biology

Centring academic workflows on focus and memorization

I believe that one of the most significant challenges for academics in relation to digital information consumption is how to gain control over the content they intend to explore digitally. This is particularly true of literature monitoring, for academics cannot afford to placidly ignore alerts, podcasts and advanced online publications if they want to stay competitive, but cannot (or should not) let themselves become overwhelmed by a constant flow of literature to read. 'Tuning' the flow of information that digital platforms provide is thus essential to achieving the mindfulness and focus required for scholarly activity. The idea that the omnipresence of digital devices reshapes our way of thinking and memorizing is not new, but should be taken seriously when thinking of the user-information relationship. Indeed, the way we interact with digital environments has deep implications for the ability of modern libraries, public repositories, journal article websites and reference management software to respond to the needs of their end users.

" 'Tuning' the flow of information ... is ... essential to achieving the mindfulness and focus required for scholarly activity."

Recent studies on internet browsing behaviour indicate that the attention span of internet users is in decline. In 2013, it was estimated to be eight seconds on average, which embarrassingly makes average internet users less well performing than goldfish⁴. This ever-shortening attention span has many consequences for avid internet users, especially for the youngest generations. It highlights the (in)ability to balance multitasking and focus appropriately, especially when it comes to using digital devices.

How do 'impatient' internet users read, memorize and learn? It seems that the answer is "not so well". For instance, studies indicate that far less than one third of words present on an average-sized text get read when these texts are read within a web browser⁵. In other words, the longer and the less readable texts are, the quicker we lose our patience and switch to irrelevant content that disrupts concentration. The so-called digital generation does not have the monopoly on this behaviour, though: for instance, office workers check their e-mail inbox 30 times per hour on average⁶. I could pretend that thousands of hours spent reading paper-based material in a distraction-free environment (study rooms and libraries) made me immune to these trends, but from the very moment I started my PhD, I joyfully jumped into the trap of multitasking, making me too impatient to read and learn efficiently on a digital device (at first). However, I suspect that my own spontaneous browsing patterns are not different from those of most academics, who themselves are not much different from most internet users.

This suggests that publishers and app developers should account for the terrible attention spans of their users. Fortunately, it seems that some already do, since efforts to make digital information consumption more fluid have become visible in recent years. For instance, websites like ScienceDirect, PLOS or Cell Press, among others, have redesigned the presentation of online papers, while reducing the global information content of pages and making relevant information more visible. Others, like eLife or PubMed's PubReader, propose single-page web applications that rearrange the content of the traditional full-HTML manuscript to offer a more appealing experience of reading papers on a screen. The general tendency, it seems to me, is to follow the direction that modern mobile apps took: minimalist design, emphasis on richly-interactive presentation, and personalized content. The impact of design, however, has its own limitations: any given website or software is ultimately 'competing' with e-mails, social media, apps, documents and other potential distractions that coexist on the same digital device(s), and there is little that the designer can do about this.

"... average internet users [are] less well performing than goldfish."

While it might seem absurd to study in a library accompanied by chatty friends, movies, games, e-mails and music, this is exactly what happens in an uncontrolled digital environment. Even in the absence of distraction, looking at multiple monographs on the same device is in itself often disruptive. This is where the user's mindset comes into play. I consider the identification and the annotation of papers as being two very distinct tasks,

129 which require very distinct behaviours: on the one hand, I have a fast-paced and exploratory mindset when identifying new papers, but on the other hand, I need to feel slower and more focused when reading papers of interest. In most cases, I operate such a change of mindset by changing the media (e.g. from website to reference manager [RM]), the platform (e.g. from desktop computer to tablet), or the 'context' I am working within (e.g. switching off e-mails for a long period of time).

In my experience, achieving focus and memorization within an efficient academic workflow requires the combination of 1) clear design that brings content clearly and quickly to my attention, 2) clear definitions of what tools and what platforms work best for which task and 3) control over external distractions. This is what, in my opinion, allows me to 'channel' flows of information, i.e. to adapt the quantity and the relevance of information that I wish to process depending on the task I wish to achieve.

The basic components of academic workflows

In their simplest form, my academic workflows are composed of three successive stages, which I refer to as 'identify – digest – use' (Figure 1). Each stage necessitates a certain number of actions, for which specific digital tools prove to be extremely useful. I will refer to these as 'academic productivity tools' (APTs). Here, I describe examples of my workflows, highlighting both the functions I need for each task, and the APTs that I presently use that have these functions.

The first stage, 'identifying' relevant literature, can be done actively while querying catalogues or repositories with a specific question in mind. However, exploiting alternative approaches can be helpful, since the interest here is to maximize the capacity to discover relevant literature as soon as it becomes available, with a minimum of effort. The easiest way to achieve this is to delegate the monitoring to an automaton. I personally use RSS feeds, and specific alerts (such as F1000 or Google Scholar) inform me instantly when a paper with specific keywords or metadata has been detected on public repositories. I also use a cross-platform RSS aggregator (Feedly) that allows me to access all this content at one location, from my smartphone, tablet or desktop computer. Recommendation engines from publishers' websites and reference managers also help me identifying a body of literature once a paper of interest has caught my attention. In addition, my colleagues and I share the URLs of papers of interest by e-mail, which is another way to (passively) discover relevant literature.

"...exploiting alternative approaches can be helpful ..."

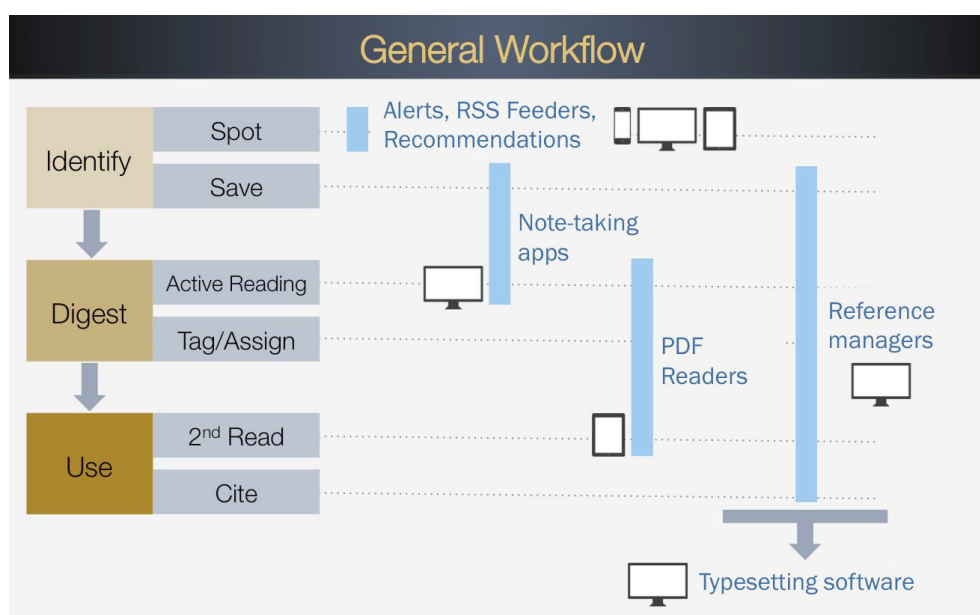


Figure 1. The basic architecture of an academic workflow

130 Accessing the article's URL from my own digital devices is advantageous, since I can directly integrate papers of interest into my workflow. I do this by saving a local copy of the full-length article PDF in a cloud-synchronized repository, which itself is monitored by an RM. I personally use Mendeley, but performant RMs exist for most platforms (in particular Papers, ReadCube, EndNote and Zotero), and all offer quite straightforward means to save references (one-click web importers, 'drag & drop' features, built-in search engines). However, one advantage of Mendeley is that it can also automatically extract metadata information whenever a PDF is saved in a 'watched folder'. In this manner, I can directly ensure that a newly saved PDF is at the same time integrated in my RM and accessible from PDF viewers and note-taking apps that I prefer to use on my tablet (e.g. GoodReader). Alternatively, I can also use recommendations that Mendeley gives, based on my own library, to discover papers of interest.

"... I can directly integrate papers of interest into my workflow."

The second stage of the workflow consists of 'digesting' information, which involves an annotation step (consisting of reading the entire paper and annotating the PDF), as well as an assignment step (where the goal is to define how the paper relates to my research projects, using classification and assigning tags). Note that I make sure to 'identify' and 'digest' papers at different times in order to maintain the right mindset for each task.

The third and final step concerns the 'use' of the paper, which consists in most cases of retrieving, reading, and citing references when preparing a manuscript or presentation. This stage might happen minutes, weeks or years after the 'digest' stage, and thus requires information to be easily retrievable from the RM. Finally, the citation itself is a basic function of all RMs. The most popular RMs seem to have adopted by default a 'cite-as-you-write' approach, which allows users to cite papers without leaving their typesetting software. This approach is flexible, as it enables you to retrieve citations directly by querying anything you might remember about the paper you are looking for: author names, date, bits of the title, bits of the abstract, journal, etc. One limitation of RMs is the current lack of compatibility with alternatives to Microsoft Word and LibreOffice Writer. Recent years have seen the increased popularity of distraction-free writing apps, as well as powerful tools like Google Docs and Scrivener, that present unique features which might appeal to academics. However, to date, using any of these alternatives presents a frustrating citing experience that only tech-savvy users are likely to accept.

The next steps

So far, I have only considered the case of peer-reviewed articles. What about the rest? In my research, I often need technical information (e.g. about programming or statistics) that I might find in specialized forums, hear during talks, discuss with experts, read in hard-copy versions of textbooks. How can such disparate sources of information be integrated in a digital academic workflow? To my knowledge, there is no obvious and easy way to achieve this via the use of an RM. However, the cloud provides a simple means to synchronize and save multimedia 'notes'. I typically use Evernote, and invest time assigning tags to ensure that all the notes I capture connect to one another and to my projects. Even though this system works, the bits of information stored within Evernote and within an RM remain independent, which means that nothing else but the user's memory guarantees that the relevant content in both resources will be accessed and used during the writing process.

"I make sure to 'identify' and 'digest' papers at different times in order to maintain the right mindset ..."

Scrivener, a cross-platform typesetting software, proposes an efficient solution for ensuring that users have control over what information they want to use during writing. One of the basic concepts of Scrivener is to allow the embedding of notes, hyperlinks, PDFs, text files, figures, spreadsheets, presentations, etc. to specific portions of text (e.g. paragraphs, sections, chapters). For instance, one can start writing the third paragraph of an introduction and read within the app the content that was linked to that specific paragraph. I believe that this approach to text composition (which grants users a great degree of freedom to bring

131 'context' to any part of a manuscript and rearrange the outline intuitively) can be of great benefit for academics writing long manuscripts. More than the choice of a specific tool, what matters here is that the features of the software subtly align with the mindset of the writer, creating a rather effortless path from the 'digest' to the 'use' part of the workflow. Nonetheless, Scrivener cannot be considered as a substitute to a note aggregator like Evernote. Using it thus creates a third 'channel' where the user stores relevant information: an RM for references, Evernote for contextualized content that might be retrieved at any point, Scrivener for content needed while writing. I guess this highlights that there is room for substantially better academic software to integrate information in a smarter way. I imagine the ideal academic software as a hybrid between a cross-platform advanced note-taking aggregator and an RM that would offer means to export citations in various writing apps, as well as embed functions for text composition, for instance by making it possible to organize independent notes together and to compile them all into a manuscript. This would, in effect, collapse the three 'channels' into one.

"...the power that digital information holds is far from being fully exploited."

In short, the power that digital information holds is far from being fully exploited. Much remains to be done for digital devices to adapt to us (rather than the opposite), and this means that technology needs to account for our lack of consistency, changes of mood and mindset. Improving user interfaces, cloud synchronization, machine learning-based recommendation engines and maximizing interoperability are the key ingredients that successful academic software should build on. In that scenario, I suspect that the PDF format is likely to decline, but its decline will remain slow until richly interactive HTML-based articles become as easy as PDFs to attach to e-mails, save and read in single-purpose apps such as advanced 'PDF' readers and reference managers. However, in the long term, the adoption of HTML-based articles as a standard will catalyze the emergence of the 'ideal' academic software I described, as HTML allows fast and easy extraction of data for apps to smartly interact with the content, and offers great ways to integrate videos and interactive content that can profoundly improve our experience of reading scientific articles.

What roles for the library?

Interestingly, the description of my academic workflows barely mentions how I interact with university libraries, and neither did my presentation at UKSG⁷. This is in part because using local library catalogues is simply less rational than using online tools that provide better, faster, more integrated solutions for search and discovery. I guess that there is little that can be (or should be) done to go against the natural tendency of users to follow paths of least resistance. In fact, I want to operate from my own devices. The reason that I actually can is that the Library at my Institute has understood how users like me want to work. Internet connections from my Institute directly grant access to all the content the Library has subscribed to, and off-site access to e-resources is made possible via a proxy. In other words, the Library allows its users to access most library services effortlessly, without having to leave their desk – sometimes without even knowing it. I believe that this is a requisite for building efficient academic workflows.

"...using local library catalogues is simply less rational than using online tools ..."

Although libraries will undoubtedly remain an essential part of academic life, it is clear that their relationship to end users has to change. But it's not bad news. Innovative, forward-thinking librarians are already changing how libraries interact with students and researchers⁸. Providing support and guidance for digital tools, training in academic workflows, improving delivery with proxies for students at home and for alumni, increasing the visibility of their institute's research on the internet: these are just a few of the missions that librarians have to tackle in the current digital environment. These are exciting times.

References

1. SINTEF, 'Big Data, for better or worse: 90% of world's data generated over last two years', ScienceDaily, 22 May 2013. <http://www.sciencedaily.com/releases/2013/05/130522085217.htm> . (accessed 2 May 2014).
2. <http://jasonpriem.org/2010/10/medline-literature-growth-chart/> (accessed 1 May 2014).
3. NIH: http://www.nlm.nih.gov/services/pubmed_searches.html (accessed 1 May 2014).
4. <http://www.statisticbrain.com/attention-span-statistics/> (accessed 1 May 2014).
5. Weinreich, H, Obendorf, H, Herder, E and Mayer, M, Not Quite the Average: An Empirical Study of Web Use, *ACM Transactions on the Web*, 2008, 2(1).
6. Weinreich, H, et al, ref. 5.
7. Chalancon, G, Channelling information flows: a young researcher's approach to knowledge management, 2014, figshare. DOI: [dx.doi.org/10.6084/m9.figshare.1004470](https://doi.org/10.6084/m9.figshare.1004470)
8. Kortekaas, S, Thinking the Unthinkable: doing away with the library catalogue, UKSG, 2014, <https://www.youtube.com/watch?v=a6BPclajLVI> (accessed 1 May 2014).

Article copyright: © 2014 Guilhem Chalancon. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use and distribution provided the original author and source are credited.



Guilhem Chalancon, MSc (Hons), Gates Cambridge Scholar, PhD Student
University of Cambridge / MRC Laboratory of Molecular Biology
Regulatory Genomics & Systems Biology, MRC Laboratory of Molecular Biology
1 Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK
Tel: +44 (0)1223 26 7831 | E-mail: guilhem.chalancon@gmail.com | web:
<http://www.mrc-lmb.cam.ac.uk/genomes/guilhem>

ORCID iD: <http://orcid.org/0000-0001-5117-6602>

To cite this article:

Chalancon, G, Channelling information flows: a young researcher's approach to knowledge management, *Insights*, 2014, 27(2), 127–132; DOI: <http://dx.doi.org/10.1629/2048-7754.151>