IJDC | *General Article*

Encouraging and Facilitating Laboratory Scientists to Curate at Source

Cerys Willoughby
University of Southampton

Jeremy G. Frey
University of Southampton

Abstract

Computers and computation have become essential to scientific activity and significant amounts of data are now captured digitally or even "born digital". Consequently, there is more and more incentive to capture the full experiment records using digital tools, such as Electronic Laboratory Notebooks (ELNs), to enable the effective linking and publication of experiment design and methods with the digital data that is generated as a result. Inclusion of metadata for experiment records helps with providing access, effective curation, improving search, and providing context, and further enables effective sharing, collaboration, and reuse.

Regrettably, just providing researchers with the facility to add metadata to their experiment records does not mean that they will make use of it, or if they do, that the metadata they add will be relevant and useful. Our research has clearly indicated that researchers need support and tools to encourage them to create effective metadata. Tools, such as ELNs, provide an opportunity to encourage researchers to curate their records during their creation, but can also add extra value, by making use of the metadata that is generated to provide capabilities for research management and Open Science that extend far beyond what is possible with paper notebooks.

The Southampton Chemical Information group, has, for over fifteen years, investigated the use of the Web and other tools for the collection, curation, dissemination, reuse, and exploitation of scientific data and information. As part of this activity we have developed a number of ELNs, but a primary concern has been how best to ensure that the future development of such tools is both usable and useful to researchers and their communities, with a focus on curation at source. In this paper, we describe a number of user research and user studies to help answer questions about how our community makes use of tools and how we can better facilitate the capture and curation of experiment records and the related resources.

Received 23 February 2017 ~ *Accepted* 7 September 2017

Correspondence should be addressed to Cerys Willoughby, School of Chemistry, University of Southampton, Southampton, SO17 1BJ. Email: cerys.willoughby@soton.ac.uk

An earlier version of this paper was presented at the 12th International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

In science, communication is of the utmost importance. Ideas need to be shared, evidence disseminated, plans discussed, findings recorded, and errors corrected. An individual researcher may work alone, but their research is of little value to the scientific community if it remains inside their heads. The increasing dominance of technology, digitization of data, and the drive towards more openness in science, brings with it a need to ensure that scientific records and their associated data are effectively captured, managed, and preserved; and, where appropriate, made accessible to others. Two essential elements of adapting to this changing environment are the effective curation of digital records and the expanding use of ELNs (Bird, Willoughby, Coles and Frey, 2013; Bird, Willoughby and Frey, 2013). ELNs provide a range of features that can help to improve data management, data retrieval, and collaboration, providing positive benefits for research, including the curation of records for collaboration and reuse (Badiola et al., 2015).

Curation is often viewed as a retrospective activity in the academic publication lifecycle, typically undertaken by someone other than the original author of the research, frequently leading to high costs, introduction of errors, and the loss of both data and context (Beagrie et al., 2010; Borgman, 2012; Downing, et al., 2008; Edwards et al., 2011; Qin et al., 2016). However, in most small and medium-sized laboratories, the scientists who produce the data are solely responsible for their management (Borgman et al., 2016), but difficulties occur in encouraging researchers to effectively curate their data for the benefit of unknown future users, particularly as the researchers have a limited understanding of the value of adding metadata and there are a lack of perceived benefits for doing so (Borgman, 2008; Crystal and Greenberg, 2005; Currier et al., 2004; Edwards et al., 2011; Frey, 2008; Greenberg, 2004). Machine-generation of metadata can help to alleviate some of these problems by making the curation process cheaper, more efficient, consistent, and enable researchers to focus on more complex tasks (Greenberg et al., 2006). However, machine-generated metadata tend to be limited to timestamps, measurements, and instrument or software properties, and so the production of 'user-defined' metadata by the expert author of the research is often necessary to capture the full context and add meaning to the data (Bird, Willoughby, Coles and Frey, 2013; Borgman et al., 2016; Currier, et al., 2004; Greenberg and Robertson, 2002; Frey, 2008; Willoughby et al., 2015). The burden of curation can be eased, if curation is designed into experiments and captured at source (Bird, Willoughby and Frey, 2013; Frey, 2008). Researchers need to appreciate that it is good laboratory practice for data creators to anticipate what researchers might encounter in later stages, by collecting, curating, and preparing for citation as they create the data – 'at source' (Frey, 2008). User-friendly tools need to be developed that not only support researchers in curating their work for the benefit of others, but also by providing functions that enable the authors themselves to gain value from the metadata generated as a result.

The Southampton Chemical Information Group has been involved in e-Science projects for over 15 years exploring the exploitation, collection, curation, dissemination, and reuse of scientific data and information (Frey et al., 2015). Comb*e*Chem (Frey, et al., 2006) instigated a number of e-Science projects, including the Smart Tea and More Tea projects designed to support chemists in the preparation, execution, analysis, and dissemination of their work (Frey, 2008; Frey et al., 2006; schraefel et al., 2004). The project conducted user studies using the analogy of ‘making tea’ to explore the experimental and recording practices of synthetic chemists to develop a prototype for an electronic replacement of the traditional paper notebook (schraefel et al., 2004). The two-part prototype included a web component, enabling chemists to create a plan for their experiment based on the assessment form used as part of health and safety requirements, stored in RDF format. The second component was a wireless tablet PC with a stylus to capture the experiment in the lab, primarily capturing planned, versus actual, measurements used, checking off steps of the experiment, and writing notes or drawing images. Researchers valued certain elements of the system such as chemical information lookup and automatic data capture, but in real-world use, the interfaces were too inflexible and prescriptive, with significant work needed to adapt the tools for other domains (Milsted et al., 2013).

Concerns about lack of flexibility led the team to investigate the opposite approach of developing a generic ELN that captured unstructured data with minimal semantics and enabled researchers from any domain to be able to capture their experiment records without constraint (Milsted et al., 2013). The resulting researcher-centric blog-based ELN, known as LabTrove, enables users to capture their research as free-form text, but also provided a way to design custom templates so that repeated experiments could be easily replicated and documented using a common format. As well as automatically capturing metadata such as author, timestamps, URIs, and version information, users are able to add their own user-defined metadata to their posts using a mixture of a mandatory single tag to describe the post (a ‘Section’), and a flexible multiple ‘key-value’ pair system (Milsted et al., 2013). The user-defined metadata, although not inherently semantic, is exploited to provide ways to search, navigate, group, and filter records within the ELN (Milsted et al., 2013). LabTrove is successfully used in a heterogeneous set of laboratories across the world, demonstrating the value of the flexible interface across a variety of laboratory science disciplines (Badiola et al., 2015).

Although LabTrove serves a number of research teams successfully, the interface was not a focus during development and remained rather crude (Badiola et al., 2015). In 2010, a series of user research activities were begun, in order to investigate current usage of the LabTrove ELN, attitudes of researchers to ELNs, what usability improvements could be made to improve the adoption of ELNs, and whether using structured record templates like Smart Tea or the ‘blank page’ of LabTrove would make a difference to the capture of experiment records.

In this paper, we provide an overview of the results of these activities with a brief background to the research, the methods that were used, a brief discussion of the major findings, and the significance of the findings in terms of designing tools and methods for encouraging curation at source for scientific researchers. We also explain the benefits of including the ability to create an experiment plan in ELNs to enable the early capture of provenance information that can support the researcher throughout their experiment. We present ideas for interface designs that can support and encourage curation for experiment records, as well as briefly discussing other methods for encouraging curation at source.

Metadata Use in the LabTrove ELN

For a full description of the LabTrove ELN, refer to Milsted et al (2013) and Badiola et al. (2015). Our user research activities began with a series of semi-structured interviews with members of the different LabTrove user communities. The purpose of the interviews was to develop an understanding of the research and practices of the different groups, find out what enhancements groups wanted to see in LabTrove, and to examine how each group was using metadata within their projects. As a result of the assessment and the discussion with the users, it became apparent that users struggled with adding metadata for their experiments and many researchers were anxious about using it. Anxieties ranged from not knowing where to start with metadata, to actively fearing the consequences of 'getting it wrong' and concern about the amount of effort that would be involved in fixing any mistakes. The lack of metadata on their entries meant they were also finding it increasingly difficult to locate specific notebook entries and details for particular experiments.

At around the same time we had the opportunity to run some user testing sessions on LabTrove with several groups of researchers at Bath University who were planning to adopt the ELN. We included metadata as one of the elements of the testing to investigate how the researchers used the metadata features within LabTrove. In these studies, we found that users added metadata only when required by the system, and often added inappropriate values, even when prompted and provided with sensible examples. Post-study discussions with subjects highlighted not only problems with the interface, but also that researchers did not have any understanding of what metadata does, what it might be for, and why it might be useful. Just asking users to add metadata in the interface was not sufficient.

To investigate whether issues of minimal metadata use were common in LabTrove, we conducted a survey of metadata use across over 100 LabTrove notebooks (Willoughby et al., 2014). As a result of the survey, we discovered that the majority of users were not really using metadata at all. A large percentage of users used a 'catch-all' term when metadata is mandated, for example a space, dash or 'General'. The majority of metadata added were high-level labels rather than specific terms; these can help inform types of metadata that would be useful in describing an experiment such as Instruments and Materials, but their lack of specificity limits the usefulness of the metadata within the system itself. There was a lack of metadata terms that described techniques or conditions in the experiment that could be useful for searching or filtering experiments. When notebooks were used by multiple authors, more metadata was used, but the privacy level of the notebooks did not affect the amount of metadata added, suggesting that users were adding metadata for both social reasons and to support personal organisation. This behaviour is in common with that observed on other platforms (Ames and Naaman, 2007; Hammond et al., 2005). Metadata was observed to be used effectively by some individuals and groups, for example a researcher using metadata in his notebook to help with structuring and writing up his thesis, and some teams where members used a common schema or shared notebooks.

After the LabTrove survey was completed, similar surveys were conducted on a variety of other platforms that make use of user-defined metadata to investigate if the pattern of metadata use within LabTrove was unusual and whether these platforms could provide an insight into how different interfaces might encourage the creation of metadata. Flickr, NASA blogs, various chemistry-related blogs, and the myExperiment website were investigated as part of the surveys (Willoughby et al., 2014). As might be expected, the metadata terms generated differed between different communities, indicating that no one set of terminology is likely to work for all, but commonalities do exist at a high level between them, particularly for high-level terms and topics. In common with LabTrove, the use of a 'catch-all' term was very common, especially where a default value was provided. In terms of effective interfaces, those that provide prompts or ‘invitations’ to use or reuse metadata appeared to be particularly effective at encouraging metadata creation and curation.

User Studies on Generating User-defined Metadata

Following from observations on the way that users used, or failed to use, metadata on LabTrove and other platforms, we conducted a study to investigate ways that different techniques might be used to generate user-defined metadata and their potential benefits and consequences. We also considered whether different techniques might be more or less successful for generating metadata for different purposes: providing additional context or by aiding retrieval by providing terms that might be used as search keywords. The studies described in this section were run in parallel, and in some cases in combination, with our studies investigating how different templates, and therefore interface designs, affected what information researchers captured about their experiments (Willoughby et al., 2016).

LabTrove, and most of the social networking platforms we investigated, collect metadata using a ‘free recall’ method where invitations are used to encourage the addition of tags, categories, or other metadata types. These invitations act as a reminder to add the metadata, but are not really acting as cues to prompt the capture of any information in particular. Cues can act as reminders for generating metadata (Crystal and Greenberg, 2005), but can also influence the information that is recalled (Higham and Tam, 2005; Marian and Neisser, 2000; Tourangeau et al., 2014). Asking someone to recall information using a different perspective, for example by describing information from the perspective of different audiences, can also change the information that is recorded (Dudukovic et al., 2004; Tversky and Marsh, 2000). Using a combination of these techniques, *Free recall*, *Cued recall*, and *Changing perspective*, is often used to maximise information collected, for example in witness interviews and questionnaires (Jobe and Mingay, 1989; Schwarz, 2007; Tourangeau et al., 2000).

Our first study investigated how *Free recall*, *Cued recall*, and *Changing perspective* might affect the generation of metadata terms using an online questionnaire (Willoughby et al., 2015). In *Free recall,* subjects were asked to write terms that come to mind when they viewed the information source; in the *Cued recall* condition, they were asked to write terms about locations, people, equipment, activities or actions, materials, and other objects; and in the *Changing perspective* condition, subjects were asked to write terms that would help someone search for a similar item. Changing the condition changed both the quantity and nature of the terms produced. *Changing perspective* generated the smallest number, most relevant, and most consistent number of terms, and also suggested sources that would be of use in a search. *Free recall* generated similar terms but a more diverse range, which often included irrelevant terms that were unlikely to be useful for curation, including personal opinion, descriptive information, and terms about things not mentioned in the content. *Cued recall* generated the largest number and broadest range of terms, matching the questions asked. A somewhat worrying observation in the *Cued recall* condition, was that important topics were significantly underrepresented, or even missing, if they were not explicitly cued. When cues were ambiguous or irrelevant, the subjects still attempted to add responses, adding terms describing personal experience, or external elements not present in the study materials.

A follow-up study presented the opportunity to try similar techniques when describing an experiment (Willoughby et al., 2015). Participants were asked to record details of an experiment involving Lego Rocket Cars using different designs of template. The templates included a *Free recall* section, or a 'Keywords' section (expected to illicit similar information to the *Changing perspective* condition above), or the same *Cued recall* questions. Similar results to the first study were observed: *Cued recall* generated the highest number, diversity, and most relevant terms; *Keyword* responses were fewer and more consistent, as seen in the *Changing perspective* condition; and *Free recall* condition generated similar terms to Keyword responses, but with some terms adding useful additional context and some less useful personal opinion.

The next study was a follow-up to our investigations into template use, enabling us to investigate metadata creation for a set of more formal experiments using *Free recall* and *Cued recall* (Willoughby et al., 2016). In this study, 20 students completed three organic chemistry experiments and were asked to record each experiment on one of three different templates, every template included a *Free recall* section, and one included a *Cued recall* section asking about chemicals, equipment, locations, techniques, and other useful information. *Free recall* was less successful at generating useful terms with the inclusion of many statements and irrelevant information, although this varied by subject. In contrast, cued responses were relevant and consistent between participants, although the naming conventions for chemicals generated by the cue for materials were diverse, with a mixture of formal chemical names, common names, and molecular formulae. A similar diversity for chemical naming was also seen in the Lego Rocket Cars study. Inconsistency in nomenclature between researchers was also raised as an issue during discussions with chemists in the ELN community. For chemicals and other topics, *Cued recall* produced terms describing the experiment that had not been included elsewhere on the template; without the *Cued recall* section, important context information would not have been captured at all.

A final study was devised to investigate the *Free recall*, *Changing perspective*, and *Cued recall* conditions for generating metadata on a much larger scale, involving 117 undergraduate chemistry students from Southampton University. As part of their first-year practical chemistry module, the students were required to complete a series of three 'remote experiments', and then complete a post-lab survey. There were three versions of the survey, one for each condition. The responses to the survey were analysed by word-type, term topics, and for relevance and personal opinion, in the same way as the first study. The results of this larger study helped to confirm the findings of the original study and were also valuable in terms of evaluating the usefulness of *Changing perspective* for generating metadata. As before, the *Cued recall* produced the largest numbers and diversity of terms, which were mostly relevant to the experiment, except for a few adjectives such as 'fun' and 'sufficient'. *Free recall* and *Changing perspective* generated similar term topics, but around a third of the *Free recall* responses were irrelevant or personal opinion, compared to the *Changing perspective* responses, which contained almost no irrelevant or personal terms. In the *Free recall* condition, almost all of the irrelevant or unhelpful responses came from statements. Details about the three experiments, questions asked on each survey, and more detailed results can be found in the Appendix.

Strategies and Interfaces for Metadata Generation

The results from the final study indicate that *Changing perspective* is effective for producing terms that would be relevant for search, reducing the amount of irrelevant information that is captured, and producing consistent values that are likely to be meaningful to different users. The study also confirmed the advantages of *Cued recall* for producing high numbers of both relevant and diverse responses for the cued categories. As discovered in the first study, however, there are two major disadvantages of *Cued recall*: if a cue is missing, important information may not be captured at all; and any ambiguity over the meaning of the cue can lead to confusion and a loss of information, or the generation of irrelevant terms. *Cued recall* can also be more laborious for the user, with multiple questions to capture the desired context, but enormous value could be gained from enabling the users to be able to customise the cues themselves (Willoughby et al., 2015). *Free recall* appears to be effective for generating a diverse range of terms that can provide context to the experiment, but also produces a lot of irrelevant and personal information that is probably not valuable for use as metadata and of little value to others trying to make use of the experiment record, especially when the responses are statements, rather than words or short phrases.

Many of the studies captured metadata at the same time as the experiment data, providing the opportunity to compare the relative effectiveness of asking the user to provide metadata, versus automatic metadata creation using data mining. In an ELN or other tool, data mining techniques can be used to extract meaning from the content of experiment records or associated data files, for example, chemical names, techniques, equipment, people, and locations. Once the terms have been mined and extracted, they can be used for other purposes in the system, such as search, navigation, filtering, grouping, and linking. For example, in LabTrove, chemical names from notebook entries can be matched and linked to records from the ChemSpider database (Day, 2013). For metadata extraction to be possible, the records must actually contain the appropriate terms in an appropriate format for them to be matched. We found, in our studies, that using separate *Cued recall* questions led to the capture of additional information that would not have been captured using the experiment narrative alone, especially for chemicals involved in the experiment. Using the responses to cued questions would be a more reliable source from which to mine metadata because more terms are generated and they are not subject to the unrecognised abbreviations, typos and ambiguous values, typically used within the report narrative.

Capturing metadata at the time of recording the experiment is extremely valuable and potentially less laborious than capturing it after the event, for example when the records are to be archived, and all the context and original details may have been long forgotten. Laboratory scientists typically create a plan to detail materials to be used and actions to be performed in the experiment, often as a part of a risk assessment process where the safety and treatment of materials need to be identified and 'signed off' before the experiment can be started. In addition to machine-generated metadata, using cued recall-type questions at the planning stage could act as useful prompts for capturing metadata during the design process, before the experiment has even begun, and even improving the quality of the plan itself. The metadata and descriptive content captured as part of a plan can be automatically added to an experiment record that is created from the plan, to minimise duplicate creation of metadata for each experiment, and to create useful links and relationships. Metadata captured in a plan can also be reused and updated when existing plans are used to develop new plans, for example to develop a technique over time, or to conduct repeating experiments, and with plans shared amongst a research group or research community. ELNs, or other tools designed to support laboratory scientists with the curation of their research, need to provide effective management of the creation and sharing of plans and risk assessments, in addition to other data such as experiment records, results, and analyses. Such management tools, and workflow and plan sharing tools such as myExperiment, can provide opportunities for effective reuse of plans by enabling the tracking of provenance, details of approvals, information about changes that have been made, and maintaining or forging new links to relevant literature and materials repositories and inventories, for repeating experiment runs, and sharing within communities.

The idea of making use of plans to capture and reuse information to support the running of experiments was exploited in the Smart Tea project, as mentioned in the Introduction. We developed the idea further by using plans to capture metadata in a proof of concept, iPad-based ELN called Notelus[1], designed for synthetic chemists. This prototype was developed based on interviewing and observing synthetic chemists in practice, and provided a notebook-like interface to capture the experiment record based on their typical recording habits. Some basic machine-generated metadata was captured for each experiment, such as timestamps and names, which the user could supplement with user-defined metadata about the experiment, such as project names, experiment conditions, descriptions, and a list of keywords. The user could also construct a simple XML-based plan, manually using an editor or using a companion iPad app, 'Plan Buddy for Notelus'[2], to add additional context and metadata to the experiment record, including a unique plan identifier, plan name and author, prepopulated experiment metadata, reaction scheme links, materials (names, descriptions, safety, molecular-weight, density, ratio, and measurements), equipment, and procedures to be used. These values were used to enhance the interface to help support the running and recording of the experiment in the lab. When the experiment record was exported to the LabTrove ELN, key-value pairs were created in LabTrove for each item of equipment and material included in the plan, and for the Plan ID, Plan Name, and Plan Author. Within the ELN, the metadata is then available for navigation, search, filtering, and grouping of the experiment records. The inclusion of metadata from the plan also ensures consistency of metadata terms and identifies the provenance of the experiment design as well as the particular experiment run and results.

Our study of metadata use in LabTrove and discussions with the ELN community indicated that simply asking a user to add metadata, without additional support, is ineffective and leads to minimal metadata use, and a lack of understanding of how to use metadata led to metadata avoidance. In the surveys of metadata use on other platforms, if users are provided with a default option for adding metadata, this default option, such as 'uncategorised' or 'general' becomes overused and useless for curation. More specific questions, or prompts about what information is desired, may encourage users to provide this information; using *Free recall*, *Cued recall*, and *Changing perspective* to consider how others might find the information, can all help. We found that many of the platforms included in the surveys, use a variety of prompts and 'invitations' to encourage users to add tags, categories or labels to their items. Invitations are a way to encourage users of software to provide information or to perform tasks. They often highlight a part of the interface and invite you to use it, by providing a link, or a button such as a plus sign, to encourage you to explore, or ask you a question to elicit particular information. Invitations act in part to attract attention, but

1 Notelus: https://itunes.apple.com/gb/app/notelus/id593269701?mt=8

2 Plan Buddy for Notelus: https://itunes.apple.com/us/app/plan-buddy-for-notelus/id663881731?mt=8

also as prompts to record information. Other social media platforms that encourage users to share information, such as LinkedIn[3] and Facebook[4], also use invitations extensively to retrieve information from users, typically in the form of invitations to 'Add' particular information, people, places, and activities, but also sometimes use questions, such as 'What's on your mind?', to elicit information.

An important element of asking the user to add metadata is to provide value from its use by displaying it prominently in the interface and using it to enable functions of value to the user. For example, in Flickr, the tags and other annotations added by the user, are visible on the photograph page and create links that display other photos with the same tags or annotations, whilst machine-generated metadata, such as automatically generated tags and photograph EXIF data, are also prominent in the interface and provide useful functionality for both the photograph owner and the audience community. Examples of this, and other example interfaces, can be found in the Appendix. A potential interface including invitations for capturing experiment metadata based on our studies is shown in Figure 1.

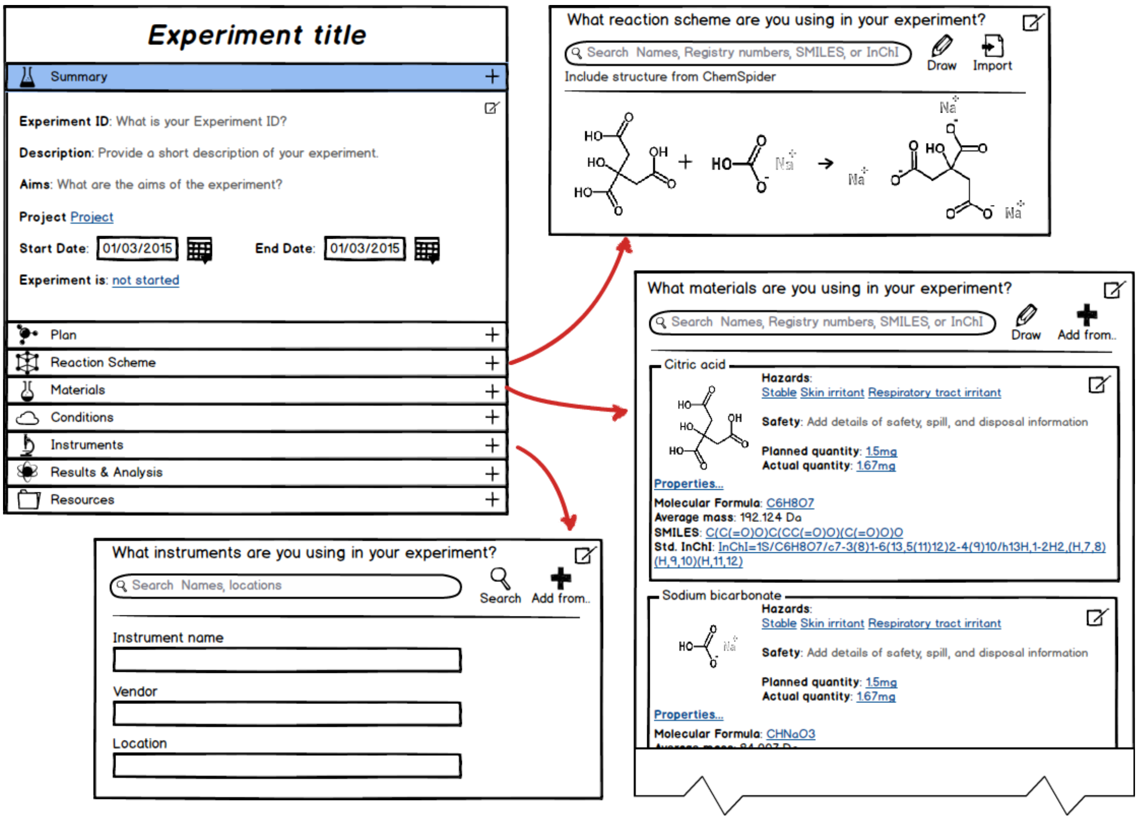


Figure 1. Mock-up of a potential interface for capturing metadata in LabTrove.

Invitations could be used in conjunction with formal metadata schemas, dictionaries, or ontologies, if these are appropriate, enabling the users to select from metadata pre-defined by a group, or that the users themselves have previously used. In addition to invitations to add metadata, it is important to also provide a way to capture information that has not been explicitly asked for and to enable customisation of the interface so that users can add their own missing prompts that are relevant for themselves and their

3 LinkedIn: https://www.linkedin.com
4 Facebook: https://www.facebook.com

collaborators, and remove or change ambiguous or superfluous prompts, so they are not 'forced' to record information that is not useful or is confusing. For example, a prompt for Materials could be duplicated and customised to generate appropriate terms depending on the discipline, such as, 'Inputs', 'Reagents', 'Strain', 'Products', 'Substance', 'Compound', 'Chemical', 'Molecule type', 'Batch', or 'Sample'. These terms are used by different disciplines to describe materials used in experiments. Some way of entering certain metadata in graphical form may also be of value, for example, in our community discussions, a desired ELN feature for synthetic chemists was the ability to search by drawing molecules rather than entering names or textual structure descriptions, a facility provided by ChemSpider[5] for example. Creating metadata by drawing the chemical structure may make it easier for the researcher to capture the chemicals and reactions relevant to their experiment, and to help others to find and retrieve their research by molecules of interest.

Customisation may also help with the tendency of metadata schemas and ontologies to change over time, as projects evolve or a better understanding of the metadata is developed (Noy and Klein, 2004). A 'one size fits all' approach does not work for metadata schemas, so the ability to design, or import, a preferred schema into a tool is important. Such tools could be designed so that importing a preferred metadata schema could automatically generate appropriate prompts within the interface or a template. As well as deriving metadata, data mining and ontologies can have value in terms of providing 'smart' assistance to users. When a user creates an experiment plan or record, chemical or equipment names could be extracted and suggested to the user as potential metadata to add, enabling the user to accept, amend, or dismiss the values as appropriate metadata for the experiment. Such systems could also make use of metadata across a team, by making suggestions based on patterns of metadata usage across a whole system, project, or community, again increasing the value of adding metadata and hopefully encouraging researchers to contribute by adding metadata to their own work.

In our LabTrove survey, and subsequent discussion with researchers in the community, we discovered that the majority do not have a good understanding of metadata and data management, matching findings from other researchers across the physical sciences and other disciplines (Goodger and Worthington, 2013; Ward et al., 2011; Rowlands et al., 2008; Whitton, 2011; Whitton and Takeda, 2011). Better data management training, creation of education resources, improved documentation, video tutorials, exemplar notebooks, metadata checklists (Higdon et al., 2014), sharing best practices, and the assistance of embedded librarians and Information Specialists (Kim et al., 2011), can all help to improve the situation, by imparting the value of metadata and the importance of curation at source.

Conclusions

Curation is often viewed as a retrospective activity in the academic publication lifecycle, perhaps not even an activity completed by the author of the research. We recognise the 'burden of curation' on researchers, but we assert that the author of the research brings the most valuable knowledge to the curation process and the development of user-friendly tools not only supports researchers to curate their work for the benefit of others, but also enables authors themselves to gain value from the metadata generated as a result. Although this paper focuses on the capture and use

5 ChemSpider Structure Search: http://www.chemspider.com/StructureSearch.aspx

within ELNs of metadata for laboratory researchers and their experiments, many of the lessons learned from the studies can help inform the design of usable interfaces to encourage effective metadata creation in other domains where researchers generate plans and conduct physical and computational experiments.

Providing users with a blank page, or relying on *Free recall* for metadata creation, encourages users to record what comes to mind, if anything, but may not yield information that is relevant to them, or others, at a later stage. In contrast, asking users to change perspective to think about how others might find and retrieve their research generates useful and consistent metadata terms. The greater variety and more personal nature of *Free recall* terms can still be effectively utilised in the interfaces, but providing ways to separate metadata for personal and social organisation would benefit the author and their audience (Ames and Naaman, 2007; Hammond et al., 2005). For the important contextual details that cannot be machine-generated, including prompts or cues in interfaces does encourage the capture of metadata. By asking the right questions, we can generate metadata and context for the experiment record that is relevant, specific, and useful, and can be reused to add value to both the authors and to future audiences of the research. The quality of metadata captured is only as good as the prompts used, and interface and template designers need to have an understanding of the needs of both the audience of the research, and the researchers themselves, to ensure important cues are not ambiguous, missing, or overwhelming. Most importantly, interfaces need to make effective use of the metadata collected so that users get active benefits from metadata they add to their records on a day-to-day basis, and not just for the benefit of an 'invisible' audience at some unknown point in the future.

Acknowledgements

The research reported in this paper was supported by Grants from Smart Research Framework (HEFCE/JISC), UMF, E-Research South (EPSRC South East Regional e-Research Consortium EP/F05811X/), and RCUK Digital Economy IT as a Utility Network+ (EP/K003569/). The authors would like to thank members of the LabTrove community, and students and researchers from University of Southampton and KTH, Stockholm, who took part in various studies. We would also like to thank Thomas Logothetis and Colin Flowers for their invaluable assistance in running the studies with students at University of Southampton.

References

Ames, M. & Naaman, M. (2007). Why we tag: Motivations for annotation in mobile and online media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*; Rosson, M. B., Gilmore, D., Eds.; Association for Computing Machinery: San Jose, CA. pp 971−980.

Badiola, K.A., Bird, C.L., Brocklesby, W.S., Casson, J., Chapman, R.T. Coles, S.J., Cronshaw, J.R., Fisher, A., Frey, J.G., Gloria, D., Grossel, M.C., Hibbert, D.B., Knight, N., Mapp, L.K., Marazzi, L., Matthews, B., Milsted, A., Minns, R.S., Mueller, K.T., Murphy, K., Parkinson, T., Quinnell, R., Robinson, J.S., Robertson, M.N., Robins, M., Springate, E., Tizzard, G., Todd, M.H., Williamson, A.E., Willoughby, C., Yang, E., & Ylioja, P.M. (2015). Experiences with a researcher-centric ELN. *Chemical Science*. doi:10.1039/C4SC02128B

Beagrie, N., Lavoie, B., & Woollard, M. (2010). *Keeping Research Data Safe 2*. JISC. Retrieved from http://www.jisc.ac.uk/publications/reports/2010/keepingresearchdatasafe2.aspx

Bird, C.L., Willoughby, C., Coles, S. J., & Frey, J.G. (2013). Data curation issues in the chemical sciences. *Information Standards Quarterly, 25*(3), 4-12. doi:10.3789/isqv25no3.2013.02

Bird, C.L., Willoughby, C., & Frey, J.G. (2013). Laboratory notebooks in the digital era: the role of ELNs in record keeping for chemistry and other sciences. *Chemical Society Reviews*, *42*(20), 8157–75. doi:10.1039/c3cs60122f

Borgman, C.L. (2008). Data, disciplines, and scholarly publishing. *Learned Publishing, 21*(1), 29-38. doi:10.1087/095315108X254476

Borgman, C.L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*. *63*(6), 1059–1078. doi:10.1002/asi.22634

Borgman, C.L., Golshan, M.S., Sands, A.E., Wallis, J.C., Cummings, R.L., Darch, P.T., & Randles, B.M. (2016) Data management in the long tail: Science, software, and service. *International Journal of Digital Curation*, *11*(1), 128-149. doi:10.2218/ijdc.v11i1.428

Crystal, A., & Greenberg, J. (2005). Usability of a metadata creation application for resource authors. *Library and Information Science Research, 27*(2), 177–189. doi:10.1016/j.lisr.2005.01.012

Currier, S., Barton, J., O'Beirne, R., & Ryan, B. (2004). Quality assurance for digital learning object repositories: Issues for the metadata creation process, *ALT-J, Research in Learning Technology, 12*(1), pp. 5-20. doi:10.1080/0968776042000211494

Day, A. (2013). Search and insert from ChemSpider in LabTrove and other websites with TinyMCE editors. Retrieved from http://www.chemspider.com/blog/search-and-add-from-chemspiderdirectly-from-labtrove-and-other-blog-based-websites-which-use-tinymce.html

Downing, J., Murray-Rust, P., Tonge, A.P., Morgan, P., Rzepa, H.S., Cotterill, F., & Harvey, M.J. (2008). SPECTRa: The deposition and validation of primary chemistry research data in digital repositories. *Journal of Chemical Information and Modeling, 48*(8), 1571–81. doi:10.1021/ci7004737

Dudukovic, N.M., Marsh, E.J., & Tversky, B. (2004). Telling a story or telling it straight: the effects of entertaining versus accurate retellings on memory. *Applied Cognitive Psychology, 18*, 125-143. doi:10.1002/acp.953

Edwards, P.N., Mayernik, M.S., Batcheller, A.L., Bowker, G.C., & Borgman, C.L. (2011). Science friction: Data, metadata, and collaboration. *Social Studies of Science*, *41*(5), 667–690. doi:10.1177/0306312711413314

Frey, J.G. (2008). Curation of laboratory experimental data as part of the overall data lifecycle. *International Journal of Digital Curation, 3*(1) 44-62. doi:10.2218/ijdc.v3i1.41

Frey, J.G., Coles, S.J., Bird, C.L., & Willoughby, C. (2015). Collection, curation, citation at source: Publication@Source 10 years on. *International Journal of Digital Curation*, *10*(2), 1–11. doi:10.2218/ijdc.v10i2.377

Frey, J., De Roure, D., Taylor, K., Essex, J., Mills, H., Zaluska, E., Moreua, L., & Foster, I. (2006) CombeChem: A case study in provenance and annotation using the Semantic Web. In L. Moreau & I. Foster (Eds.), *Lecture Notes in Computer Science*: *Vol. 4145. Provenance and Annotation of Data*. pp. 270-277. doi:10.1007/11890850_27

Goodger, J. & Worthington, W. (2013). Research data management training for the whole project lifecycle. In *Physics and Astronomy Research (RDMTPA) JISC Final Report. Research Data Toolkit.* University of Hertfordshire. Retrieved from http://research-data-toolkit.herts.ac.uk/document/rdmtpa-final-report

Greenberg, J. (2004). Metadata extraction and harvesting. *Journal of Internet Cataloging, 6*(4), 59–82. doi:10.1300/J141v06n04_05

Greenberg, J., & Robertson, W.D. (2002). Semantic web construction: An inquiry of authors' views on collaborative metadata generation. Paper presented at the *International Conference on Dublin Core and Metadata for e-Communities. Italy*. Retrieved from http://dcpapers.dublincore.org/pubs/article/view/693

Greenberg, J., Spurgin, K., & Crystal, A. (2006). Functionalities for automatic metadata generation applications: A survey of metadata experts' opinions. *International Journal of Metadata, Semantics and Ontologies*, *1*(1), 3. doi:10.1504/IJMSO.2006.008766

Hammond, T., Hannay, T., Lund, B., & Scott, J. (2005). Social bookmarking tools (I): A general review. *D-Lib Magazine,* 11. Retrieved from http://www.dlib.org/dlib/april05/hammond/04hammond.html

Higdon, R., Stewart, E., Stanberry, L., Haynes, W., Choiniere, J., Montague, E., … Kolker, E. (2014). MOPED enables discoveries through consistently processed proteomics data. *Journal of Proteome Research*, *13*(1), 107–13. doi:10.1021/pr400884c

Higham, P.A., & Tam, H. (2005). Generation failure: Estimating metacognition in cued recall. *Journal of Memory and Language*, *52*(4), 595-617. doi:10.1016/j.jml.2005.01.015

Jobe, J., & Mingay, D. (1989). Cognitive research improves questionnaires. *American Journal of Public Health, 79*(8), 1053–1055. doi:10.2105/AJPH.79.8.1053

Kim, Y., Addom, B., & Stanton, J. (2011). Education for eScience professionals: Integrating data curation and cyberinfrastructure. *International Journal of Digital Curation*, *6*(1), 125–138. doi:10.2218/ijdc.v6i1.177

Marian, V. & Neisser, U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology General, 129*(3), 361–368. doi:10.1037/0096-3445.129.3.361

Milsted, A., Hale, J., Frey, J., & Neylon, C. (2013). LabTrove: A lightweight, web based, laboratory "blog" as a route towards a marked up record of work in a bioscience research laboratory. *PLoS ONE*, *8*(7), e67460. doi:10.1371/journal.pone.0067460

Noy, N. & Klein, M. (2004) Ontology evolution: Not the same as schema evolution. *Knowledge and Information Systems*, *6*(4), 428–440. doi:10.1007/s10115-003-0137-2

Qin, J., Dobreski, B., & Brown, D. (2016) Metadata and reproducibility: A case study of gravitational wave research data management. *International Journal of Digital Curation*, *11*(1), 218–231. doi:10.2218/ijdc.v11i1.399

Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., Gunter, B., Withey, R., Jamali, H.R., Dobrowolski, T., Tenopir, C. (2008). The Google generation: The information behaviour of the researcher of the future. *Aslib Proceedings*. *60*, 290−310. doi:10.1108/00012530810887953

schraefel, m.c., Hughes, G.V., Mills, H.R., Smith, G., Payne, T.R., & Frey, J. (2004). Breaking the book. In *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04* (pp. 25–32). New York, USA: ACM Press. doi:10.1145/985692.985696

Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology, 21*(2), 277–287. doi:10.1002/acp.1340

Tourangeau, R., Conrad, F.G., Couper, M.P., & Ye, C. (2014). The effects of providing examples in survey questions. *Public Opinion Quarterly, 78*(1), 100–125. doi:10.1093/poq/nft083

Tourangeau, R., Rips, L.J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.

Tversky, B., & Marsh, E.J. (2000). Biased retellings of events yield biased memories *Cognitive Psychology, 40*(1), 1-38. doi:10.1006/cogp.1999.0720

Ward, C., Freiman, L., Jones, S., Molloy, L., & Snow, K. (2011). Making sense: Talking data management with researchers. *International Journal of Digital Curation 6*(1), 265–273. doi:10.2218/ijdc.v6i2.202

Whitton, M. (2011). *Institutional data management blueprint (IDMB) questionnaire quantitative data*. University of Southampton: Southampton. Retrieved from http://eprints.soton.ac.uk/195959

Whitton, M. &Takeda, K. (2011). *Data management questionnaire results: IDMB project.* University of Southampton: Southampton. Retrieved from http://eprints.soton.ac.uk/196243/1/IDMB_Survey_Report.pdf

Willoughby, C., Bird, C.L., Coles, S.J., & Frey, J.G. (2014). Creating context for the experiment record. User-defined metadata: Investigations into metadata usage in the LabTrove ELN. *Journal of Chemical Information and Modeling, 54*(12), 3268–83. doi:10.1021/ci500469f

Willoughby, C., Bird, C.L., & Frey, J.G. (2015). User-defined metadata: Using cues and changing perspectives. *International Journal of Digital Curation*, *10*(1), 18-47. doi:10.2218/ijdc.v10i1.343

Willoughby, C., Logothetis, T.A., & Frey, J.G. (2016). Effects of using structured templates for recalling chemistry experiments. *Journal of cheminformatics*, *8*(1), 1. doi:10.1186/s13321-016-0118-6

Appendix 1: Remote Experiments Study

Undergraduate students from the University of Southampton were asked to complete three remote experiments:

- **Viscosity experiment**: In this experiment, a form of falling sphere apparatus was set-up and run for various liquids. Videos taken of the experiments have been set up online enabling students to make appropriate measurements to calculate the viscosity of the various liquids. Conducting the experiment in reality requires space, access to equipment, and is very messy.
- **Gas Law Experiment**: In this experiment, the students can make measurements of pressure, temperature, and water levels in the equipment set-up to investigate how the pressure and volume of a gas change at different temperatures. The equipment set-up automatically resets itself by heating the vessel at fixed times, allowing measurements to be taken in real time at any time of the day.
- **Beers Law Experiment**: In this experiment, the student can remotely control a Class 3B laser, which they would otherwise not have access to, making measurements as the beam passes through a solution in order to make calculations about the absorbance properties of the solution.

After completion of each experiment, they were asked to complete one of three post-experiment surveys to generate responses for one of three conditions: *Free recall*, *Changing perspective*, and *Cued recall*, as shown in Table 1.

Table 1. Questions asked in the three post-experiment surveys.

Survey condition	Questions asked
Free recall	Write down up to ten things that come to mind when you think about the experiment that you completed today. Write down these thoughts in the order that you think of them.
Changing perspective	Imagine you are helping someone else to find information about a similar experiment to this on the internet, what words or phrases come to mind for use in a search engine? Write as many as you can below.
Cued recall	• What list of words or short phrases would you use to describe the chemicals or other materials used in this experiment? • What list of words or short phrases would you use to describe the instruments or equipment used in this experiment? • What list of words or short phrases would you use to describe the locations used in this experiment? • What other information would you use to describe the experiment?

The responses to the surveys have been analysed by word-type, topic classification, and also for relevancy and personal opinion.

Results of the Study

As can be seen in Table 2, in common with the other studies, the greatest number of responses are generated in the *Cued recall* condition, and the fewest are generated in the *Changing perspective* condition.

Table 2. Average and total number of terms generated for each condition.

	Free recall condition	*Changing perspective* condition	*Cued recall* condition
Total number of terms generated	750	620	1043
Average number of terms generated	6.41	5.49	9.07

Word cloud visualisations of the responses for each of the experiments in each condition, can be seen in Figures 1 to 10. It can be seen that the *Changing perspective* condition has the fewest number of terms, but also the most consistent use of terminology between the conditions.

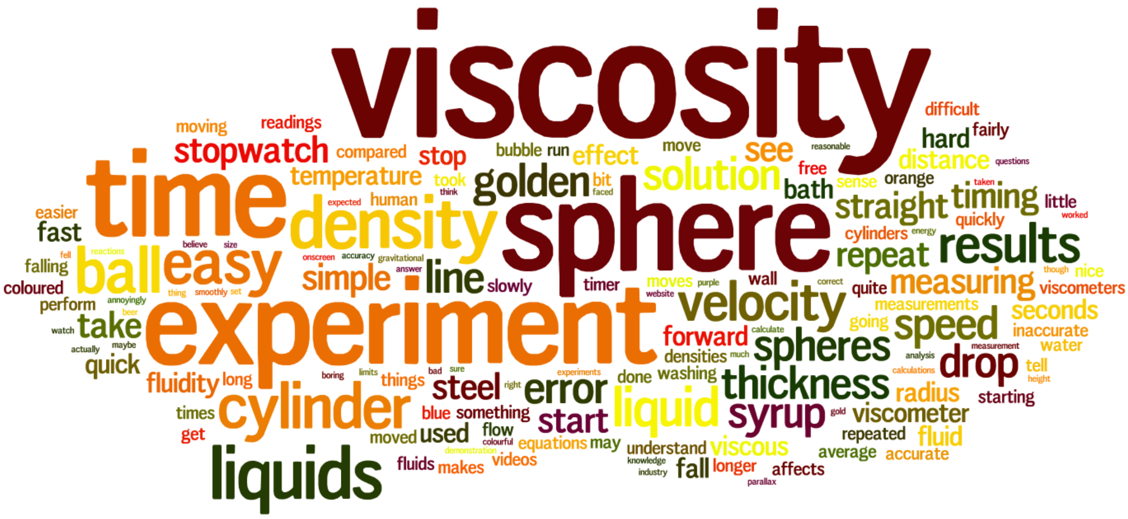


Figure 2. Word cloud visualisation of the Viscosity experiment in the *Free recall* condition.

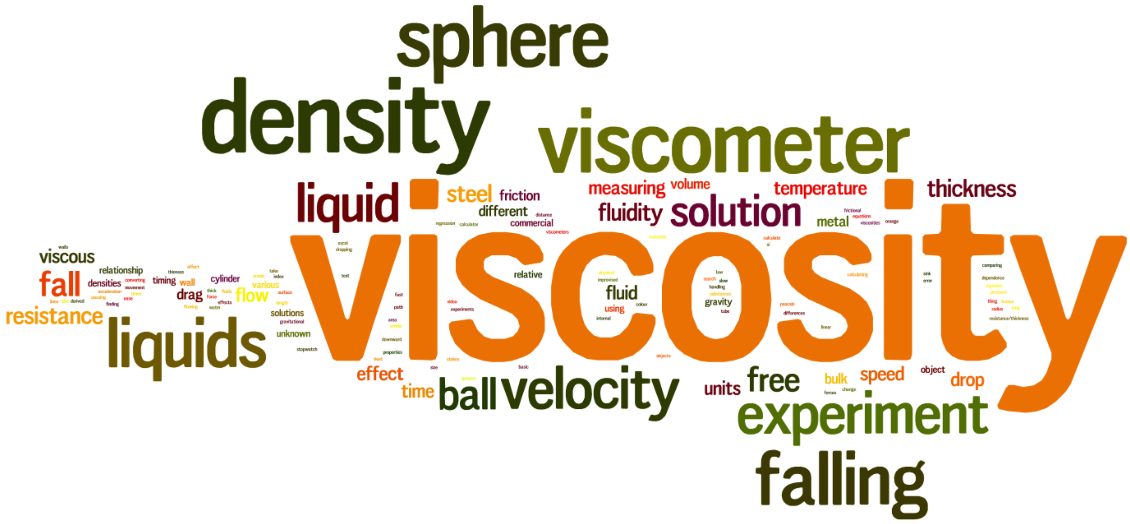


Figure 3. Word cloud visualisation of the Viscosity experiment in the *Changing perspective* condition.

Figure 4. Word cloud visualisation of the Viscosity experiment in the *Cued recall* condition.

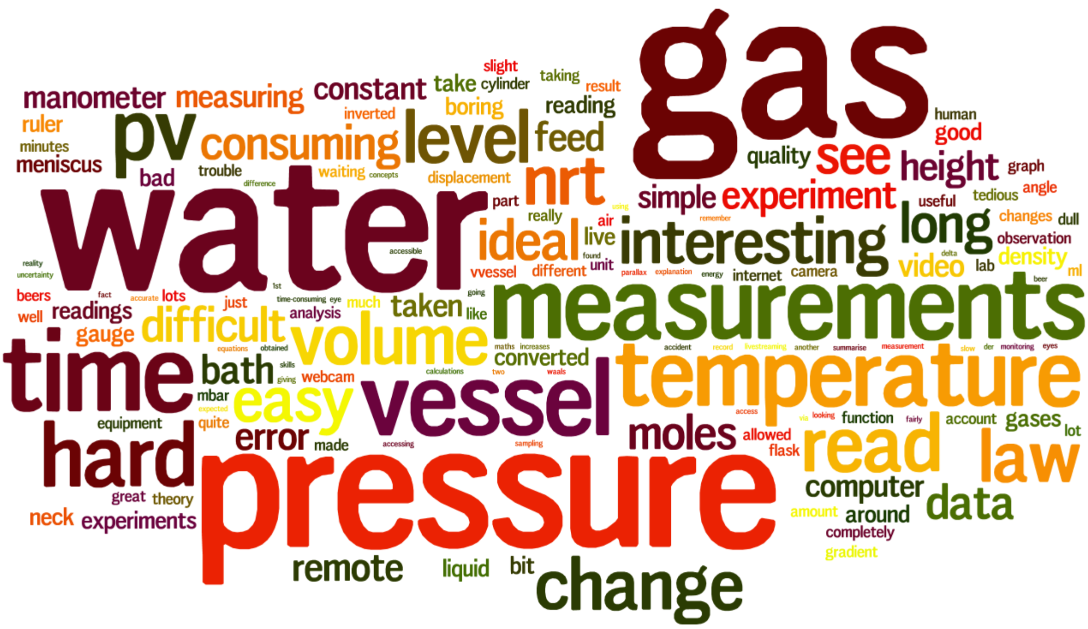


Figure 5. Word cloud visualisation of the Gas Law experiment in the *Free recall* condition.

Figure 6. Word cloud visualisation of the Gas Law experiment in the *Changing perspective* condition.

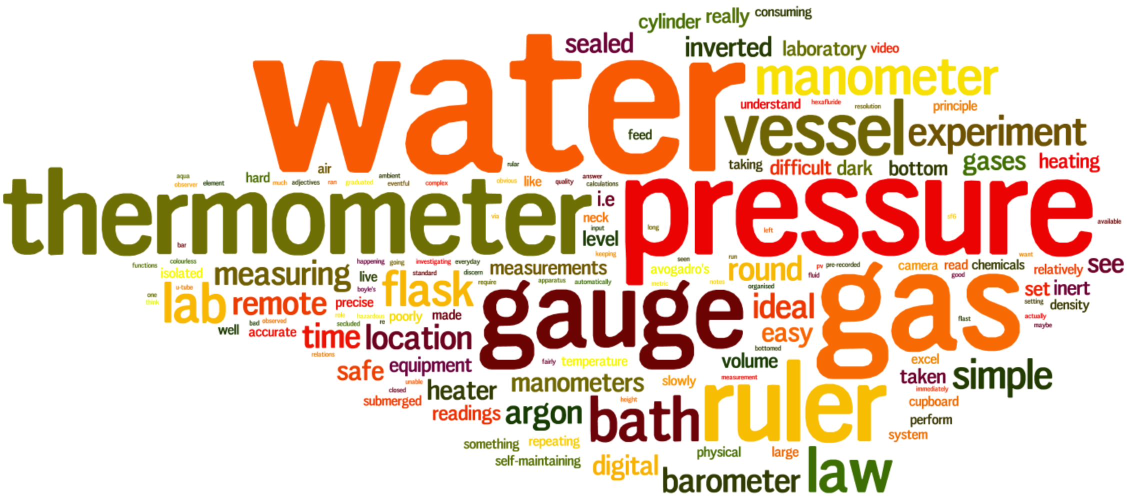


Figure 7. Word cloud visualisation of the Gas Law experiment in the *Cued recall* condition.

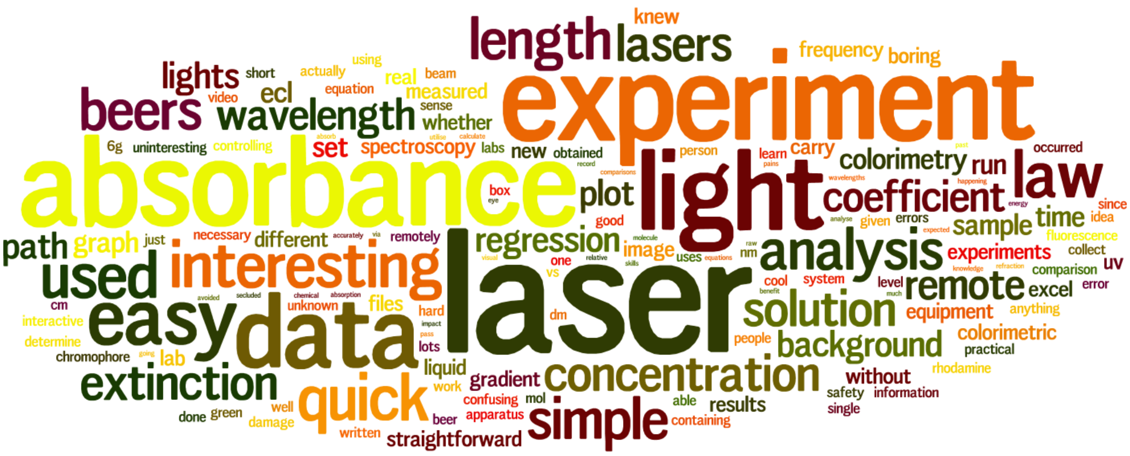


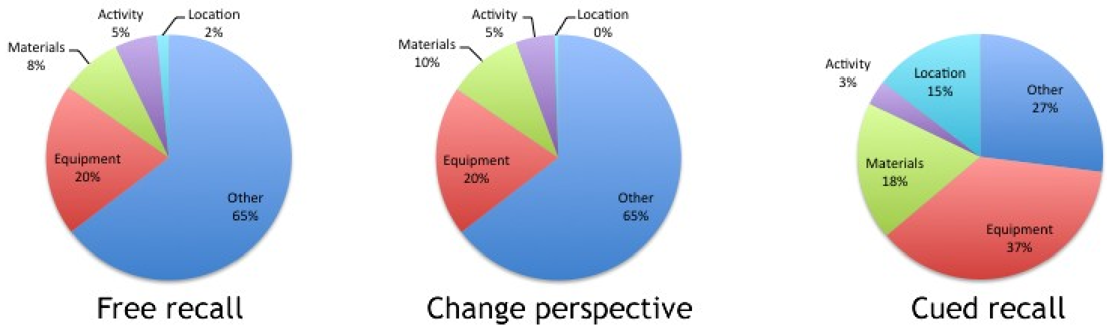
Figure 8. Word cloud visualisation of the Beers Law experiment in the *Free recall* condition.

Figure 9. Word cloud visualisation of the Beers Law experiment in the *Changing perspective* condition.

Figure 10. Word cloud visualisation of the Beers Law experiment in the *Cued recall* condition.

Across all conditions, the responses are dominated by nouns and statements, as was seen in the first metadata cues and perspectives study (Willoughby et al., 2015), although the number of adjectives is lower in the *Free recall* and *Changing perspective* conditions, and the number of verbs is reduced across all conditions. In the *Changing perspective* condition, the number of nouns is higher and with fewer statements compared to the other conditions, contributing the relevancy of the responses generated in the condition.

The topic distributions for the conditions are similar in both the *Free recall* and *Changing perspective* conditions, with responses in both conditions primarily classified as ‘Other’, followed by Equipment, Materials, and Activities, with a very small number of responses about the Location. The topics in the *Cued recall* condition are more evenly distributed, with a reduced number of responses classified as ‘Other’, and the majority of other responses matching the cues in the survey; Equipment has the highest number of responses in this condition. The full breakdowns are shown in Figure 11.



 doi:10.2218/ijdc.v12i2.514

Figure 11. Overall topic classifications for the three conditions.

The *Changing perspective* condition, contained almost no personal or irrelevant information at all, compared to the *Free recall* condition, where around a third of all of the responses are not specifically about the experiment, or represent an opinion. In the *Cued recall,* less than 10% of the responses are irrelevant to each of the cued categories; however, 50% of the responses in the 'Other information' section of the survey are personal opinion, and not particularly useful as metadata, for example: 'easy to complete' and 'boring'.

Statements are frequently used to either provide comments that are not very relevant to the experiment (e.g. 'Not sure what i'll have for dinner'), comments about them as remote experiments (e.g. 'Interesting Concept of Livestreaming the Setup' and 'The experiment was fairly convenient and quick to complete'), or comments about problems with the experiment format, particularly relating to difficulties making measurements or network connectivity. A number of statements also detailed the steps of the experiment or the theoretical background. In the *Free recall* condition, almost all of the irrelevant or unhelpful responses come from statements. In the *Cued recall* condition, adjectives such as 'fun' and 'sufficient' account for a large number of responses that were not particularly relevant to the question cue.

These results were particularly useful in terms of evaluating the usefulness of *Changing perspective* for generating metadata for experiments. For these experiments, the *Changing perspective* condition actually captured some of the most useful responses, generating the fewest number of statements, the most consistency of terms, and producing almost no irrelevant information.

Appendix 2: Example Interfaces That Encourage Metadata Creation

Invitations are a way to encourage users of software to provide information, or to perform tasks. They often highlight a part of the interface and invite you to use it, provide a link or a button such as a plus sign to encourage you to explore, or ask you a question to elicit particular information. Invitations act in part to attract attention, but also act as prompts to record information. Often social networking platforms use these kinds of invitations to capture metadata from users as demonstrated in the following examples.

Asking questions and adding metadata from Facebook (Mock-up)

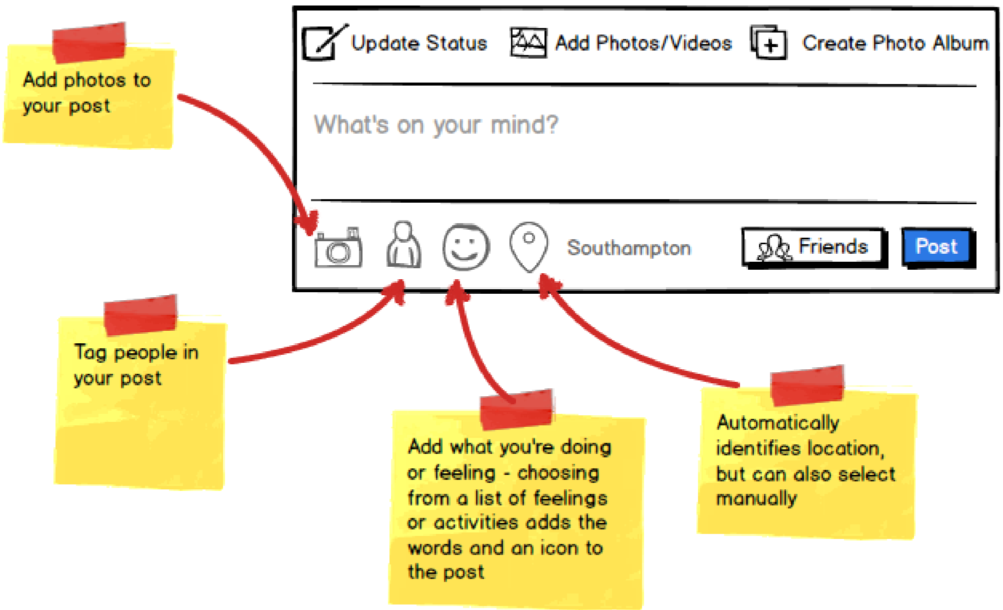


Figure 12. Asking questions and adding metadata from Facebook (Mock-up).

Figure 12 shows an invitation example from Facebook[6], where the user is asked a question about what they are thinking. Users are also encouraged to add more detailed metadata about how they are feeling, what they are doing, where they are, and who they are with. 'What's on your mind?' could be considered a prompt for free recall of information, whereas the other questions are specific and can be considered as cues.

Asking questions and tags from LinkedIn (Mock-up)

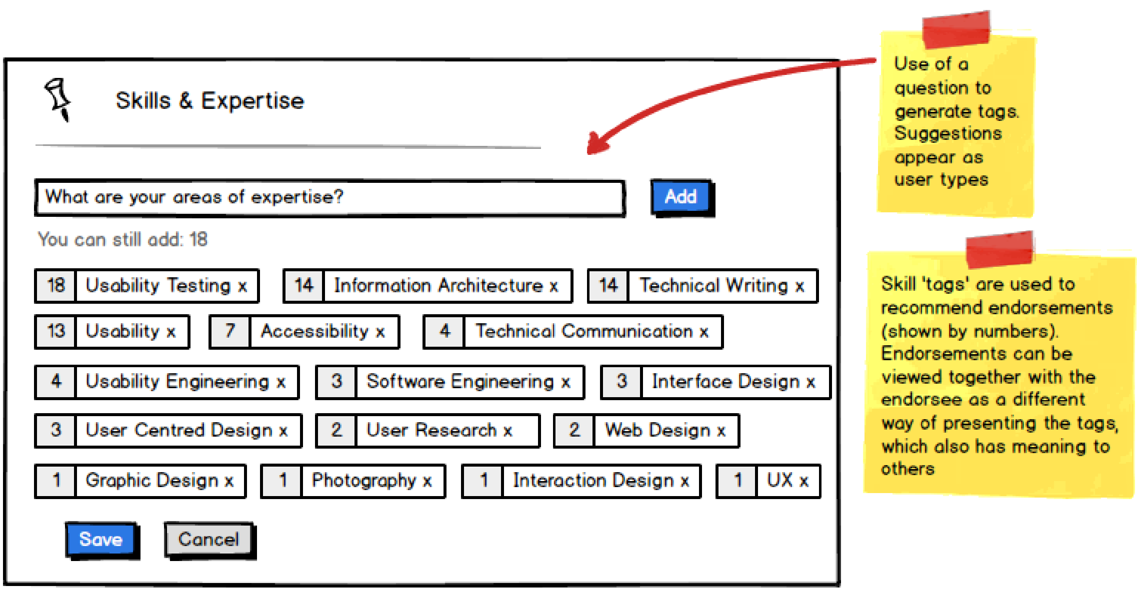


Figure 13. Asking questions and tags from LinkedIn (Mock-up).

6 Facebook: https://www.facebook.com

Figure 13 shows an example of a LinkedIn[7] invitation with a question used to elicit metadata about skills. The user is invited to add their skills in the textbox and autocomplete will match them to skills that other users have chosen. The skills that are added are displayed prominently within the editing interface shown, indicating to the user the number of endorsements they have for each skill, but are also prominently displayed for other users in their profile and included in search results, making it more likely that the user will be found and matched with appropriate contacts or career opportunities.

Invitations and metadata elicitation example from LinkedIn (Mock-up)

Metadata is also collected in many of these social networking platforms in the form of 'profiles', where users are encouraged to complete a series of fields, usually about themselves, or to describe an object. For example, Figure 14 shows a mock-up of an invitation from LinkedIn, encouraging users to add information that has not yet been included in a personal profile. Invitations include the phrase 'Recommended for you', an arrow pointing at the interface, and the 'plus' buttons. Clicking on the plus symbol reveals a form to capture the metadata for this particular content. Placeholders are left within the profile for sections without content, as a reminder to complete them.

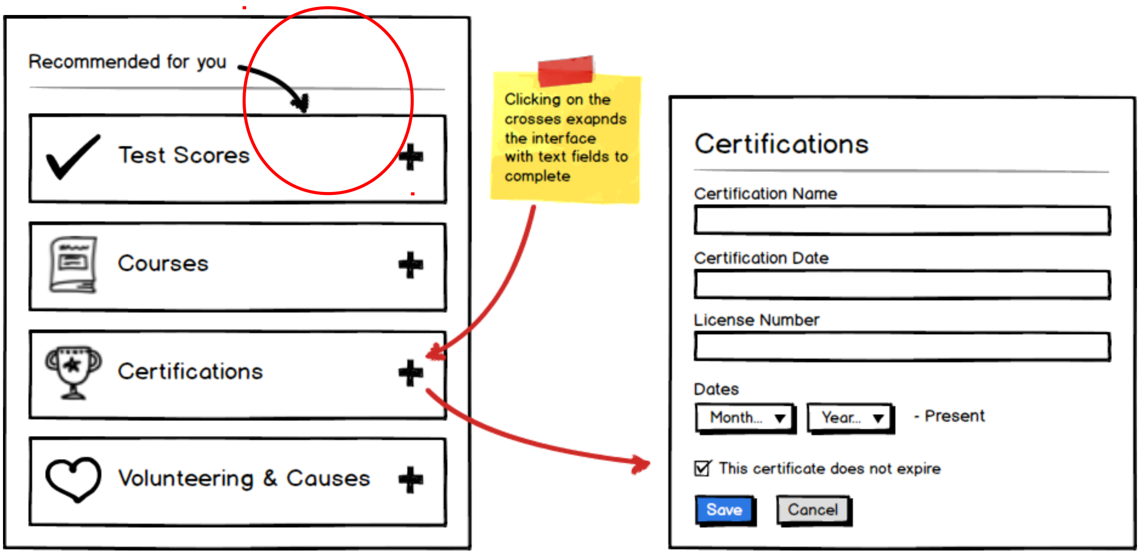


Figure 14. Invitations and metadata elicitation example from LinkedIn (Mock-up).

7 LinkedIn: https://www.linkedin.com

Photo 'Profile' from Flickr (Mock-up)

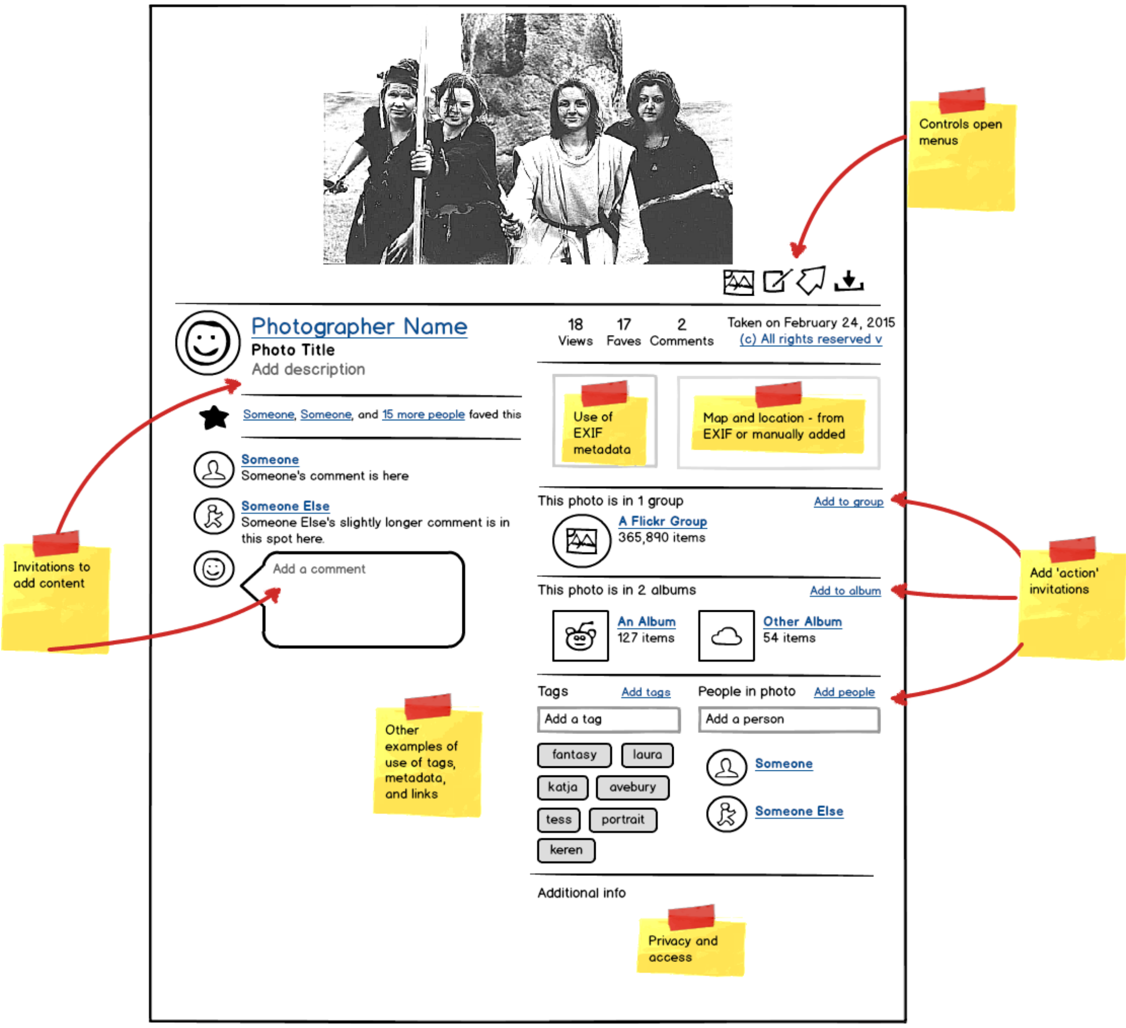


Figure 15. Photo 'Profile' from Flickr (Mock-up).

Figure 15 shows a mock-up of the Flickr photograph interface showing a 'Profile' for a photograph containing a number of invitations to perform actions and create new content. The user is encouraged to add a title, comment, tag, and add people in the photograph through placeholders with instruction text, perform actions such as add the photograph to an album or group, and through examples provided in the interface. If information is missing, such as no tags or location, invites are still visible in the interface to encourage you to add that additional metadata.

The Flickr interface makes effective use of metadata, for example EXIF data is used to display information about when the picture was taken, the camera and settings for the photo; geotags are used to display a map of where the photo was taken; statistics about views and activity are displayed; what groups, albums, and galleries the photo belongs to; people in the photograph; in addition to the tags added by the user. Most metadata are displayed as links that can be used for searching, grouping, and filtering content.