



Citation for published version:

Ball, A, Whyte, A & Donnelly, M 2010, Social Networking Tools for the DCC. Digital Curation Centre.

Publication date:
2010

[Link to publication](#)

Publisher Rights
CC BY-NC-SA

University of Bath

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Social networking tools for the DCC

Alex Ball, Angus Whyte, Martin Donnelly
UKOLN, University of Bath; University of Edinburgh; HATII,
University of Glasgow

DCC REPORT
Deliverable CD 3.1

Version: 1.0

Status: Internal Review

Date: 23rd June 2010

Copyright



© Digital Curation Centre, 2010. Licensed under Creative Commons BY-NC-SA 2.5 Scotland: <http://creativecommons.org/licenses/by-nc-sa/2.5/scotland/>

Catalogue Entry

Title	Social networking tools for the DCC
Creator	Alex Ball; Angus Whyte; Martin Donnelly (author)
Subject	microblogging; blogging; social networking; forums; wikis; collaborative writing; social bookmarking; social media
Description	The DCC is considering how it can engage more fully with its target communities using social media. Encouraged by the experience of other organisations in this area, the DCC has conducted surveys both internally and externally to gauge where its efforts could be most usefully directed. It has also reviewed social media services and tools in 10 different categories. On the basis of these activities, the DCC has put forward a proposed social networking strategy that specifies the DCC's target communities, aims, outcomes and objectives, and recommended platforms and tools; it also suggests an intellectual property policy, operating processes and points for further consideration and action.
Publisher	University of Edinburgh; UKOLN, University of Bath; HATII, University of Glasgow
Date	8th March 2010 (creation)
Type	Text
Format	Portable Document Format version 1.4
Language	English
Rights	© 2010 UKOLN, University of Bath; University of Edinburgh; HATII, University of Glasgow

Citation Guidelines

Alex Ball, Angus Whyte, Martin Donnelly. (2010). *Social networking tools for the DCC* (version 1.0). Edinburgh, UK: Digital Curation Centre.

Contents

1	Introduction	4
2	Previous work	6
2.1	DigitalPreservationEurope	6
2.2	Jadu survey	8
3	Survey of DCC communities	9
3.1	Blogging and microblogging	9
3.2	Discussion and networking	10
3.3	Social bookmarking	10
3.4	Sharing literature	10
3.5	General points	11
4	Survey of DCC Staff	12
5	Review of social networking services	15
5.1	Microblogging	15
5.2	Blogging	17
5.3	Discussion and networking	20
5.4	Cross-posting, aggregation and monitoring	23
5.5	Collaborative writing	25
5.6	Social bookmarking	27
5.7	Sharing references to literature	30
5.8	Sharing images	34
5.9	Sharing presentations	35
5.10	Sharing video	36
6	Proposed social networking strategy	39
6.1	Target communities	39
6.2	Aims, outcomes and measurable objectives	40
6.3	Communication patterns	42
6.4	Recommended platforms and tools	42
6.5	Intellectual property policy	47
6.6	Operating processes	47
6.7	Implementation plan	48
A	Survey e-mail text	49
B	SurveyMonkey survey text	50
	Bibliography	52

I Introduction

In Phases One and Two of the DCC, the focus for community engagement was predominantly on providing events, mechanisms and forums where data creators, scientists, manager and librarians could find advice and support from each other and from DCC staff. The International Digital Curation Conference, Digital Curation 101 Workshops, Research Data Management Forum, DCC Forum and DCC Associates Mailing List are just a few of the ways in which the DCC has engaged (and continues to engage) those on the front line of the battle to curate data for present and future generations of researchers. In each case, the DCC has provided the forum and invited members of the community to participate.

Now that the DCC has started its third phase, it is taking the opportunity to refine its approach. While many of the initiatives of the first two phases have been successful, and will continue to be developed, two issues have become apparent. One is that the community has not participated in some of the initiatives as much as expected or hoped. The other is that the DCC could be more active in forums provided by others, forums in which the DCC's target communities are already active. In order to address these issues, the DCC has been looking at the ways in which social media platforms could be used to reach out to data creators, scientists, manager and librarians where they already are. The conclusions and recommendations from that exercise are presented in this report, in the context of the observations from which they developed.

Social networking is important to the DCC for three related reasons:

1. The fundamental reason, and the basis for this report, is the need to make more and better use of social networking to engage with our target groups.
2. Our ability to define 'digital curation'. The term is gaining popular currency, reflecting the need to distinguish valuable information from noise on the Internet. Popular definitions will influence how our stakeholders view our own role. The DCC's 'horizon scanning' role needs to monitor, if not influence, how popular definitions are evolving through their use in social media.
3. The growth of 'social curation': social networking is not just a medium for conversations about curation, it is increasingly seen as a tool for curation itself - a means to deal with the signal-to-noise ratio issue by using 'crowdsourcing' as the means of selecting to what to attend. Social curation concepts and tools are part of the requirement set for research data management. They also increase our need to advocate the broader view that curation is more than selection.

This report is primarily for internal use by the DCC, though other organisations may be interested in both the process and the conclusions of the exercise. The report begins by reviewing some previous work in this area performed by DigitalPreservationEurope and Jadu. The results of a survey conducted to gauge social media usage among the DCC's

target communities are presented in chapter 3, while chapter 4 presents the result of an internal consultation on what purposes social networking should serve for the DCC. In chapter 5, 10 different categories of social networking services and tools are reviewed in order to produce recommendations for the DCC to pursue. Finally, the conclusions of these surveys and reviews are synthesised into a proposed social networking strategy in chapter 6.

2 Previous work

2.1 DigitalPreservationEurope

The DCC is not the first digital curation organisation to undertake a formal review of what social networking services have to offer in terms of community engagement. Hutař, Melichar, Heister, Cubr and Guttenbrunner (2009) report on the usage of Web 2.0 and social networking services – both internally and externally – in DigitalPreservationEurope (DPE), and indicate how such services were integrated with the project Website and what future projects might learn from their experience.

The methodology developed by DPE was first to identify the communities it most wanted to target, and then to discover which Web 2.0 platforms are commonly used by these communities. The next step was to prepare and test a communication plan, and finally to register on the selected platforms and address the target communities on digital preservation issues.

The DPE identified as their target communities the cultural heritage sector (archives, libraries, museums, etc.), governmental and health institutions with a high reliance on digital resources, and ICT communities and companies. The genera of applications DPE investigated were blogs, discussion fora, communication tools (voice and text messaging), mapping, social bookmarking, shared calendars, shared images, collaborative writing, shared videos, social spaces (virtual worlds) and aggregation services.

Discussion. DPE set up a Facebook page as a way of building a community around digital preservation issues.¹ The page contained basic information about the project as well as links, photos, videos, presentations and publications. The aim of setting up the page was to enable interested parties to register as ‘fans’ of DPE, and thereby receive updates about project activity. Among the advantages of this over a regular Website were that additional interactivity was provided with no extra effort, and that it brought project content into a Website users were already visiting frequently. Furthermore, the updates received by fans were also visible to friends visiting those fans’ pages, allowing the information to spread wider than would be possible with, say, a standard newsfeed. DPE found that it complemented the main Website, as content there could quickly and easily be repurposed on Facebook, and Facebook increased the number of interested visitors to the DPE Website.

Collaborative writing. Internally, DPE experimented with using Google Docs as a means of collaboratively writing reports. One of the benefits of this approach was the

1. DPE’s Facebook page, URL: <http://www.facebook.com/pages/DigitalPreservationEurope-DPE/38843690994>

clear way in which individual contributions could be seen, which acted as a motivator for authors to contribute. Another was the speed and ease with which a document could be started, and the way in which versioning issues were eliminated. There were however several disadvantages identified: only basic document designs were possible, and with no pagination long documents could be difficult to navigate. Some caution was expressed over ownership of the data, and the risk of the service being withdrawn.

Externally, DPE took an active role in adding and correcting information held by Wikipedia on digital preservation issues, at both novice and expert levels. Given Wikipedia's reach and profile, DPE found this to be a worthwhile activity, but sounded several notes of caution. Promoting project outputs and information, whether directly in text or as related links or citations, must be done carefully and sensitively to avoid being considered spam. Also, Wikipedia editors take plagiarism seriously, but are not infallible when judging which text is the prior one. Finally, working with Wikipedia editors to resolve issues can be very difficult.

Social bookmarking. DPE selected Delicious as its social bookmarking service. It set up a corporate profile and used it to draw up an open list of bookmarks for DPE resources and other Web pages of relevance to digital preservation.² This list of bookmarks was integrated into both the DPE Website and the DPE Facebook page. Individuals working on the project were encouraged to tag relevant resources under their personal profiles as well as the corporate profile; this helped to raise the visibility and profile of DPE, and drive up traffic to the Website. The steady accretion of new bookmarks served to encourage other Delicious users to track what DPE was tagging.

Sharing images, presentations and documents. DPE took as one genus the services that offer to host images, presentation slides and documents. The service selected was SlideShare, which DPE used to distribute slides from training events and conferences. The facility allowing users to embed the presentations in their own Websites was particularly valued, and hosting DPE presentations under a corporate profile made it easy to move from one DPE presentation to another, and made the profile more visible.³ DPE also used Issuu to present text documents in a similar way, but found this to have much less impact; it suggested that such a service might be used more effectively for graphically designed documents, rather than standard word-processed documents.

Sharing video. DPE produced the 'Team Digital Preservation' series of animations to raise awareness about digital preservation issues among the general public. These, and subsequently some more serious live action pieces, were hosted on YouTube.⁴ While well received, there was concern about such resources getting lost in the sheer quantity of videos available. DPE also produced training videos which were hosted on the DPE Website; these too attracted positive feedback.

2. DPE's Delicious bookmarks, URL: <http://delicious.com/digitalpreservationeurope>

3. DPE's Slideshare profile, URL: <http://www.slideshare.net/DigitalPreservationEurope>

4. DPE/WePreserve's YouTube channel, URL: <http://www.youtube.com/user/wepreserve>

Conclusions. The lessons learned by DPE included three key points. The first was that project staff should be encouraged to promote the project through their own personal social networking activities. The second is that projects should measure their social networking activity against specific goals, and ensure their communications respond to user expectations. Lastly, projects should consult with professional Web marketing companies on marketing materials to be used in a social networking context. DPE also stressed the usefulness of collaboration tools such as Skype and wikis in the context of distributed projects.

2.2 Jadu survey

In 2009, CMS vendor Jadu conducted a survey of social media usage, challenges and benefits within UK Higher Education Institutions (HEIs) (Jadu, 2010). A SurveyMonkey questionnaire was used, and advertised using Twitter, blogs and a mailing list; it attracted 60 responses from 44 HEIs.

The survey found that building a business case was seen as the most significant challenge for HEIs adopting social media, with overcoming cultural issues seen as a significant challenge for adoption to date, and responding to growing user demand a significant challenge for the future. The types of social media used most heavily within HEIs were blogs and online forums, while the types of social media most frequently used for external engagement were Twitter, video sharing services such as YouTube, and networking tools such as Facebook and MySpace. Customised services such as Ning (social networking) and Yammer (microblogging) had relatively low levels of usage compared to more general platforms. The same pattern emerged when respondents were asked about their plans for future adoption of social media. They reported only marginal improvements to areas such as corporate reputation and profile, access to information and student recruitment through the use of social media, but expected the improvements to become more significant over time.

3 Survey of DCC communities

In order to determine which social media platforms are most popular among members of the communities served by the DCC, a simple survey was conducted. The survey was conducted on two platforms in parallel. An e-mail was sent to the Research Data Managers and DCC Associates mailing lists;¹ this attracted five responses. At the same time, an online questionnaire was set up using SurveyMonkey and publicised on Twitter and the Digital Curation Blog; this attracted ten responses. The text of the e-mail is reproduced in Appendix A while the SurveyMonkey questionnaire text is reproduced in Appendix B.

The spread of roles played by respondents was as in the table below. Note that respondents could specify more than one role.

Role	Frequency
Data Manager/Curator	3
Data Centre/Archive/Repository	3
IT Services	1
Library Service	5
Researcher	4
Social Media Officer	1

The e-mail and the questionnaire each received a null response, with one saying they just used a traditional website to disseminate content, and another saying they have a Twitter account they do not use.

3.1 Blogging and microblogging

Six of the respondents used both blogging and microblogging tools, while three blogged but did not microblog, three microblogged but did not blog, and one did neither. Unsurprisingly, all the microbloggers used Twitter, though two also used Facebook and one also used Identi.ca. A variety of blogging platforms were in use, the most popular being WordPress (four plus one aspiration) and Blogger (three). A few respondents mentioned services they used for tracking blogging activity, such as FriendFeed and Technorati.

1. Research Data Managers JISCMail list archive, URL: <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=research-dataman>; DCC Associates mailing list archive, URL: <http://www.mail-archive.com/dcc-associates@lists.ed.ac.uk/>

There was some variety in the reasons people cited for reading and writing blogs and microblogs. The most common reason was as a way of connecting with peers, closely followed by keeping up to date with new resources and initiatives, and monitoring current issues. Two respondents valued the ability to get quick answers to questions. Others saw blogs and microblogs as a good way of getting a feel for unfamiliar problem spaces, as a useful event management tool, as a tool for collecting relevant information for others to consume and repurpose in various ways, as an efficient dissemination tool, and as a source of unofficial information and candid opinions.

3.2 Discussion and networking

There was no sense of gravitation towards a single platform for discussion and networking among the respondents. The platforms mentioned more than once were mailing lists, FriendFeed and Facebook. Others included LinkedIn, Academia.edu and Ning. Several respondents quoted more private methods of communication and collaboration, such as content management systems, wikis, voice-over-IP systems, instant messaging and Google Wave. Points of contact with the other categories in the survey were also apparent, with mentions of Twitter, Delicious and Yahoo Buzz.

Again, the reasons for using these platforms include getting expert answers and opinions, engaging with peers, linking ideas, and monitoring the activity within a field. DCC mailing lists and FriendFeed came particularly recommended, and while one respondent was impressed with Ning they were less impressed with its charging plans.

3.3 Social bookmarking

The most popular platform for sharing links to Web resources was Delicious (with one respondent cautiously recommending the 'data_curation_resources' tag), closely followed by Twitter and Facebook. More personal media such as e-mail and instant messaging were also used. One respondent has set up a cascade so that links can be sent from Google Reader to Delicious and posted automatically on blogs, Facebook and Twitter.

The aspects of social bookmarking most valued by respondents were ease (and speed) of use, trustworthy recommendations, cloud-based storage of bookmarks, and having a venue for discussing interesting resources. There were mixed feelings about how Delicious organises links: while the tags are useful, some felt they do not provide enough structure.

3.4 Sharing literature

For the most part, the respondents did not cite specialist tools for sharing references to papers and reports. The most popular platform, used by six of the respondents, was

Twitter. Two used Delicious, two used project wikis, while others used FriendFeed, instant messaging and the first-generation Web solution of tracking the Web pages of individual academics. The most popular specialist tool, used by three respondents, was Mendeley, although one respondent did not associate it with this category of tools; the other specialist platform was Academia.edu.

The question of the benefits of sharing references, or of particular platforms, only attracted three responses. One respondent again valued having a source of high quality literature recommendations, while another valued the timeliness of such recommendations, and the sense of community they foster. The third response praised the control of permissions and formatting available in Confluence wiki software, but admitted that maintaining public-facing wiki pages can be a daunting task.

3.5 General points

The survey concentrated on some key types of social media for the sake of brevity, and so did not enquire about, for example, sharing videos or presentation slideshows. It was nevertheless interesting that one respondent volunteered their appreciation for scholarly content on YouTube, SlideShare and Prezi (alongside some other, less social sources).

One of the aspects of social media in general that respondents valued was the ability to manipulate the information in various ways, particularly where that information is made available through a newsfeed. One respondent uses Yahoo Pipes to aggregate feeds of news and resources into a single stream for perusal in Google Reader. Another filters a set of feeds for items of particular interest in FeedDemon.

One respondent summed up their usage of social media as follows.

Overall, I'm using tools for three purposes: to ask questions/get help from the community; stay aware of what's going on; and to inform my local community and others about what I and my institution are working on and thinking about.

4 Survey of DCC Staff

DCC staff were consulted on what purposes social networking should serve for the DCC. The consultation took the form of an online survey. Questions and responses are below.

Question 1 *Whose input do we need to make social networking work for us, and how?*

This is partly a question of which target groups you think we should be targeting, e.g. researchers, research administrators, funding body staff, data managers/curators, Data Centre/Archive/Repository managers, Library/IT Service managers. There may be others whose input you think would be valuable, e.g. people coordinating social media use in our partner organisations.

I think a number of social network tools can be effective for us even if they are one-way (and without 'input' from others in the accepted sense). Those tools can be used to reach all of the audiences identified in the question, and at different times we'll want to do that. Tools that do involve a real dialogue or exchange will necessarily have to target a smaller audience, as the Phase 3 plan already recognises that we can't reach everyone individually.

I guess intermediaries, as that seems to be the main audience we're targeting in phase 3, so research support, lab & IT technicians, research offices, data centres, repository managers. . . We should consult these groups and base decisions on what they're likely to use, when and how.

It would be good to look at successful models elsewhere and see what has worked for them too, especially from the digital curation field. I've noted some lessons from a social network I've recently set up below but it's not d[igital]p[reservation]-related.

We've just set up a social network for our allotment site and did a phased release which worked very well. A few committee members tried it out at first, then we passed it on to 4–5 more web-savvy plotters before putting together a basic 'how-to' guide and sending registration details to all members. It meant there was quite a bit of content across the site and a number of active members to show momentum and help get others started.

Members of the DCC; researchers engaged with the business of data management/curation; data custodians (repository, library, IT etc. staff); PR/communications units in HEIs and agencies supporting HEIs (JISC, RIN, etc.).

Question 2 *What kinds of shared interests and needs should be addressed?*

The question is not after a definitive topic list, just to establish what topics you think we should realistically expect to ‘engage’ on, because there are existing communities or platforms they fit to. For example are there any that fit closely to our web structure – events, resources, training, etc? Or what about specific themes like research integrity, open science, electronic lab notebooks?

Promoting our wares – events, resources, training – is definitely one area, and that’s one where two-way input isn’t necessary. Areas such as research integrity are more about promoting discussion or sampling views. Sometimes we might be doing this as prelude to creating resources, training or events, sometimes as an end in itself.

Social means of doing this also help us identify new people with an interest in what we’re doing, who might then become part of that ‘shadow DCC’ you’ve spoken of.

I guess this would work best by reflecting on current issues/news stories, or playing devil’s advocate to encourage debate. Some topics will get more debate than others, so I guess we should steer clear of things that won’t prompt people to add comments e.g. event announcements.

Let’s keep it at a high(ish) level: events, projects, resources and training in the fields of research data management, open science and e-science, information policy.

Question 3 *What value will be provided by sharing knowledge and experience?*

How in your view can our efforts add value to what is already going on out there?

Value for us and our target communities in promoting awareness and use of what we do. Value for us in engaging in discussion in a way that minimises use of our time, particularly if we can spark debate between others and (at times) merely be catalysts for it and observers.

Since part of our job is to identify useful things done by others and promote them, this is ‘adding value’ to those things.

I think this is precisely what we should be doing. There is so much information out there (new reports, tools, events, etc.) that it can be overwhelming and difficult for people to keep abreast of issues or navigate to what actually matters. It would be good for the DCC site to be a voice of authority, synthesising and reflecting on all this content to help users find what's relevant to them.

I think it would be good for staff to do more reflective, opinion-style pieces on what's happening and emergent themes they're spotting so we're seen as having a finger on the pulse. We already have this knowledge and skill but users don't know so don't always think to come to us. Being a bit more upfront could gain us recognition and make people more likely to engage with us for advice/consultancy etc.

By functioning as a respected national exchange or clearing-house we increase the incidence repetition of correspondents engaging with these myriad and diverse topics.

Question 4 *How is recognition or reward given for ongoing cooperation?*

Are there ways of attributing and recognise contributions that have value? For example, are there mechanisms like Technorati's Authority ratings that we should use?

Some mechanisms have their own built-in values. Commenting on a highly-rated blog, with a link back to your own, improves the rating and visibility of your own blog, for instance. Sometimes, simply acknowledging assistance (such as a tweet or retweet about us) is useful. Producing event outputs can work as well – look at what Eduserv do after events that they run; this often includes things like wordclouds of event tweets that indirectly highlight and recognise those who made lots of contributions. It's a crude measure, but people do look at it, and even compete to do well in it.

Yes, it would be good to have authority ratings. A simple sidebar with active members/biggest contributors etc. could help too. It would be good to split DCC and external users if we did that, so we're not the main ones to keep showing up.

It isn't. If you are talking about recognition of the DCC's impact (which somehow we must more accurately record), it will be made manifest by further funding.

5 Review of social networking services

In each of the following sections, we look at a different type of social media platform and consider which if any of the available alternatives would be an appropriate platform for the DCC to target. We consider in particular the impact and tone of each platform, and how comfortable or intuitive it is to use. We consider both human and machine interfaces to the platform, and how the DCC might be able to monitor its usage of the platform both from the perspective of justifying the effort spent on it, and preserving the useful outcomes of such engagement.

5.1 Microblogging

Microblogging is a medium of expression characterised by brevity, with each entry approximately the size of an SMS text message. This makes it unsuitable for expressing complex sentiments, but ideal for rapidly disseminating key information such as breaking news headlines or changes to the state of something (e.g. the current speaker at an event). While dedicated microblogging platforms exist, some other platforms have a microblogging component embedded within them, though typically with a lesser sense of narrative connection between entries, e.g. Facebook statuses, Skype profiles.

Since microblogging entries are so brief, syntax has developed to help concentrate meaning; for example, hashtags for indicating subject matter, and short codes for indicating a quote. Dedicated microblogging platforms tend to provide facilities based on this syntax, such as transforming hashtags into links to search for other entries containing the same hashtag. URL shortening services such as bit.ly and TwitPwr are used extensively, for obvious reasons.¹

Twitter. Twitter is by far the most popular microblogging tool, with around two billion entries (tweets) produced monthly at the time of writing.² It is widely used in the digital curation community for announcing resources, publicising events, providing commentary on them, and so on. Indeed, many events now specify a hashtag to use to demarcate a Twitter back-channel, and anecdotal evidence suggests that both delegates and interested non-delegates find this useful. Conversations are possible using a special syntax: '@username' for a public reply and 'd username' for a private, direct message; however, since such conversation tends to be rather ephemeral, some special effort is

1. Bit.ly Website, URL: <http://bit.ly/>; TwitPwr Website, URL: <http://twitpwr.com/>

2. Twitter Website, URL: <http://twitter.com>

needed to archive it, and the nature of medium means that as discussions get deeper they need to be taken off into a more verbose medium.

The Twitter API provides a full machine interface to the platform, and a substantial ecosystem of tools has grown up around it as a result. As well as general tools for making it easier to use the service, such as Tweetdeck³ or the Twitter iPhone and Android applications, several specialist tools are available. BirdHerd allows many individuals to contribute to a single, corporate Twitter account.⁴ Mentions of keywords (not just hashtags) may be monitored in real time using Monitter, or if traffic is low via e-mail alerts using TweetBeep; the real-time results of hashtag searches can also be embedded in Web pages, allowing Websites to 'follow' selected subjects or communities.⁵ Twapper Keeper can be used to build an archive of tweets that either include a specific hashtag, a specific keyword, or a specific username;⁶ a more comprehensive archive is planned by the Library of Congress (Shiels, 2010), allowing the current restrictions on search results (7–10 days, or 1500 tweets) to be lifted. Furthermore, some other social networking platforms now incorporate twitter, including FriendFeed and LinkedIn.

There are, however, problems that have yet to be overcome; Twitter lacks an intuitive way of following conversations, and a way of contextualising tweets without resorting to large sequences of hashtags. As with other communication media, there are issues of scalability, as following too many active Twitter users (twits) can result in information overload.

There are some alternative microblogging services very similar to Twitter, most significantly identi.ca, Jaiku and Qaiku.⁷ Identi.ca is notable for being more interoperable than Twitter in terms of the protocols it supports, while Qaiku has better multilingual support. None of these services have the penetration of Twitter, however.

Yammer. Yammer provides a sort of private Twitter, in as much as users can only see entries posted by others from the same e-mail domain.⁸ Users can link their Twitter and Yammer accounts such that any tweet of theirs containing the '#yam' hashtag is automatically echoed on Yammer. The service is free to use, but enterprises may purchase additional security and admin rights for their network.

Given that DCC staff retain institutional e-mail addresses rather than adopt dcc.ac.uk addresses, it is unlikely that Yammer could be used effectively as an internal tool, and by its nature could not be used for outreach.

Google Buzz. Google Buzz is a microblogging service that can be plugged into Google Mail.⁹ As well providing 'status update' functionality, it allows content to be imported

3. Tweetdeck Website, URL: <http://www.tweetdeck.com/>

4. BirdHerd Website, URL: <http://birdherd.com/>

5. Monitter Website, URL: <http://monitter.com/>; TweetBeep Website, URL: <http://tweetbeep.com/>

6. Twapper Keeper Website, URL: <http://twapperkeeper.com/>

7. Identi.ca Website, URL: <http://identi.ca/>; Jaiku Website, URL: <http://www.jaiku.com/>; Qaiku Website, URL: <http://www.qaiku.com/>

8. Yammer Website, URL: <http://www.yammer.com>

9. Google Buzz Web page, URL: <http://www.google.com/buzz>

from Twitter, Picasa, Flickr, Google Reader and YouTube. The service was involved in a privacy scandal shortly after launch, and it is unclear how many people now use it.

Recommendations. As Twitter provides by far the biggest reach, it is recommended that the DCC concentrate its microblogging efforts on that platform. In order to maintain brand identity, it is recommended that a corporate account be used in conjunction with BirdHerd.

5.2 Blogging

Unlike microblogging, blogging allows time and space for reasonably complex ideas to be put forward. It tends to favour writing with a ‘voice’ and a point of view: faceless corporate blogs are rare. Even though blogs seem at first glance to be a broadcast medium, they are in fact a social one, in part because the personal perspectives expressed lend themselves to cross-commentary, and in part because of the comments that can be attached directly to blog posts. These comments are an intrinsic part of most blogs; the best blogs manage to create a ‘conversation’ amongst commenters.

5.2.1 Blogging software

WordPress. WordPress is probably the most widely used blogging software.¹⁰ It is an open source project with a large user community. Blogs may be hosted on the WordPress Website, or on installations of the software on a local Web server. WordPress supports collaborative blogging, with multiple authors able to contribute to a single blog. The software comes with a user registration system with support for profiles. It is possible to set for each individual blog post whether it is public or private, and to give permission to view to specific users only.

WordPress supports W3C HTML standards and the pingback (de facto) standard for tracking links to a post. It has an in-built facility for automatically blocking comment spam, a published API and a wide range of plug-ins. The feeds generated by the software are optimised for indexing by Technorati (see section 5.2.3).

Blogger/Blogspot. Blogger is also very widely used, and since it is owned by Google it is probably sustainable for the foreseeable future.¹¹ It is currently used to host the Digital Curation Blog, and was used in Phase 2 to host the DCC Blawg and RDMF blog. Blogger is very straightforward to use and well regarded – many Blogger blogs have a high Technorati Authority rating.

Blogger supports the pingback (de facto) standard for tracking links to a post. It has a published API and a wide range of plug-ins, but is not as function-rich as WordPress.

10. WordPress Website, URL: <http://wordpress.org/>

11. Blogger Website, URL: <http://www.blogger.com/>

Recommendations. Given that the DCC's new Website includes support for blogging, any proposal to use an alternative platform would need to be supported by a strong argument outweighing those of efficiency and infrastructural simplicity. Experience with the Website so far has not revealed any noticeable deficiencies in functionality or usability; the only substantial obstacle is the fact that the Digital Curation Blog already has an established presence on Blogspot.¹² It is therefore recommended that the DCC Website be used to host any future blogging activity. The DCC should also investigate whether it would be possible to populate the Blogspot version of the Digital Curation Blog with posts from the version on the DCC Website.

5.2.2 Subject blogs

As mentioned above, blogging is a social activity, and in order to properly engage with its communities, the DCC should not only post its own material but also read and post comments on other relevant blogs. The DCC should also write blog posts that respond to or repost relevant entries from other blogs and forums, both to help draw attention to them and to attract trackback links. Below are some platforms that host curation-related or science-related blogs, upon which the DCC could comment where and when they relate to DCC activities. Blog searches (see subsection 5.2.3) would also be worthwhile to identify relevant blogs.

Nature.com Blogs. Blogs is a blog tracking and indexing service hosted by Nature Publishing Group.¹³ It collects together blogs by Nature editors and journalists, members of Nature Network and selected third-party science bloggers.

ScienceBlogs. ScienceBlogs is run by Seed Media Group, and hosts over eighty blogs dedicated to science.¹⁴ Among the potentially relevant blogs on this site are *Confessions of a Science Librarian*, Dorothy Salo's *Book of Trogoool*, and *Christina's LIS Rant*.¹⁵ There are also scientists blogging on their field – e.g. *Discovering Biology in a Digital World*, on bioinformatics¹⁶ – and many more general 'citizen science' blogs.

New Scientist blogs. *New Scientist* publishes four collaborative blogs run by staffers on the magazine, carrying posts from them and invited participants, and commenting on general science-related issues. All four are potentially relevant:

- *Short Sharp Science*: 'cutting edge science cut up'.¹⁷

12. Digital Curation Blog, URL: <http://digitalcuration.blogspot.com/>

13. Nature.com Blogs Website, URL: <http://blogs.nature.com/>

14. ScienceBlogs Website, URL: <http://scienceblogs.com/>

15. *Confessions of a Science Librarian* blog, URL: <http://scienceblogs.com/confessions/>; *Book of Trogoool* blog, URL: <http://scienceblogs.com/bookoftrogoool/>; *Christina's LIS Rant* blog, URL: <http://scienceblogs.com/christinaslisrant/>

16. *Discovering Biology in a Digital World* blog, URL: <http://scienceblogs.com/digitalbio/>

17. *Short Sharp Science* blog, URL: <http://www.newscientist.com/blogs/shortsharpscience/>

- *CultureLab*: ‘where books, art and science collide’.¹⁸
- *The S Word*: ‘politics of science and vice versa’.¹⁹
- *Big Wide World*: ‘life after graduation’.²⁰

OpenWetWare. Catering mainly to biology researchers, OpenWetWare is ‘an effort to promote the sharing of information, know-how, and wisdom among researchers and groups who are working in biology and biological engineering’.²¹ It promotes Open Notebook Science and has a blogging section, currently with 16 blogs including Cameron Neylon’s *Science in the Open*.²²

5.2.3 Blog search engines

Blogs can be a useful resource for discovering the latest developments in a field, and therefore horizon-scanning activities can be enhanced by using dedicated blog search engines.

Technorati. Technorati rates blogs and provides other feedback and social mechanisms.²³ The ‘Technorati Authority’ rating is a well accepted rating from 0 to 1000 of a blog’s ‘influence within its subject category’. A site’s ‘Technorati Rank’ is its rank among the Authority ratings across all sites, or within a topic category (‘Topical Rank’). The rating method is not disclosed. Technorati describe it as ‘calculated based on a site’s linking behavior, categorization and other associated data over a short, finite period of time. A site’s authority may rapidly rise and fall depending on what the blogosphere is discussing at the moment, and how often a site produces content being referenced by other sites’ (‘Technorati Authority FAQ’, n.d.). Authority rating is very sensitive to activity. So on the ranking scale (0–1000) the DCC Website’s Blogs section has a ranking of 0 at the time of writing, while the Blogspot version of the Digital Curation Blog has a ranking of 1, having rapidly declined from around 500 due to inactivity.

Google Blog Search. Google Blog Search has surpassed Technorati as the most popular blog search engine.²⁴ Its search algorithm is even more opaque than Technorati’s and produces quite different results, so if a search is worth doing it is probably worth doing on both.

18. *CultureLab* blog, URL: <http://www.newscientist.com/blogs/culturelab/>

19. *The S Word* blog, URL: <http://www.newscientist.com/blogs/thesword/>

20. *Big Wide World* blog, URL: <http://www.newscientist.com/blogs/bigwideworld/>

21. OpenWetWare Website, URL: <http://openwetware.org/>

22. *Science in the Open* blog, URL: <http://cameronneylon.net/category/blog/>

23. Technorati Website, URL: <http://technorati.com/>

24. Google Blogsearch, URL: <http://blogsearch.google.co.uk/>

5.3 Discussion and networking

While this report has taken a wide definition of social networking, there are services that conform to a much narrower definition. Such services enable users to build up a network of 'friends' with whom they hold discussions, share content, collaborate and compete in games, and so on.

Facebook. Facebook reportedly overtook MySpace as the world's leading social networking site in mid-2008 (McCarthy, 2008), since when its lead has increased to a comfortable margin.²⁵ More generally, it has been in the top five most visited websites for several years. Users are given a 'wall' on which they and their friends post content, typically 'status updates' and comments thereon. The contents of the wall are also available as a newsfeed. Functionality has increased over the years – the abilities to tag 'likes' and to add comments, for example, were adopted from FriendFeed (q.v.) – and may be further augmented by add-on applications such as a discussion forum. Of particular note is a Facebook to Twitter application that allows Facebook users to post tweets to their Facebook profile or Facebook page, and to find and follow Facebook friends who tweet.²⁶

Facebook is increasingly being used by education institutions to communicate with students, at least in the US (Joly, 2009). It has also been used by curation-related projects for dissemination, e.g. DigitalPreservationEurope (DPE) and CASPAR.²⁷ Having said that, it is primarily seen (and used) as a personal networking tool rather than for serious academic use. Facebook's handling of privacy settings is an ongoing issue for personal users, and may have a negative impact on its popularity.

From a DCC perspective, Facebook is a serious contender as tool for maintaining contact with the Associates Network and other users. It provides a mechanism for promoting the DCC to 'friends of friends' and would enable the DCC to collect demographic information on connections made using the service.

Nature Network. Nature Network is a social networking site aimed at scientists, provided by Nature Publishing Group.²⁸ As such it has a clear affinity with the DCC's target audiences. It carries scientific blogs and aggregates others, including the Digital Curation Blog (so the DCC already has a presence there). The site carries topical discussions in various forums or alternatively on discussion boards attached to groups, which may be private or public, and to which members can post publications. Popular forums include 'Science Policy in the UK'; there is a Data Scientists group, though with only two members currently. The other notable feature of the site is Q & A – a question answering service.

25. Facebook Website, URL: <http://www.facebook.com/>

26. Facebook to Twitter application page (requires Facebook account to access), URL: <http://apps.facebook.com/twitter/>

27. DPE's Facebook page, URL: <http://www.facebook.com/pages/DigitalPreservationEurope-DPE/38843690994>; CASPAR's Facebook page, URL: <http://www.facebook.com/group.php?gid=156672624598&ref=ts>

28. Nature Network Website, URL: <http://network.nature.com/>

Blogging aside, it is not yet clear which services will be sustainable and to which it would be worth the effort to contribute. The forums and groups overlap and have few members, and it is difficult to see how the Q & A feature will evolve as questions vary from the very general ('What are your favourite popular science books?') to the specific ('How to assay DGAT activity using a spectrophotometer?'). While the facility for posting publications to groups is novel, it is unlikely to compete with dedicated reference-sharing tools (see section 5.7).

LinkedIn. LinkedIn describes itself as a 'professional' network.²⁹ According to Wikipedia, 'the purpose of the site is to allow registered users to maintain a list of contact details of people they know and trust in business. The people in the list are called *Connections*. Users can invite anyone (whether a site user or not) to become a connection' ('LinkedIn', 2010). This list of connections can then be used in several ways, for example as a source information to find people, job vacancies or business opportunities. LinkedIn not only allows one to take advantage of one's direct connections, but also one's indirect connections up to the third degree (i.e. a connection of a connection of a connection); if there is someone among one's indirect connections one wishes to know, the mutual direct connections offer a route to gain an introduction. This 'gated access' approach is intended to foster trust among users of the service.

LinkedIn hosts LinkedIn Groups, which typically require signup. Groups offer discussion, news and jobs postings; while of mixed usefulness, if used well they work better than the corresponding feature of FriendFeed (see section 5.4.3), for example. This facility is already being used by projects relevant to the DCC, such as DataCite. LinkedIn also integrates with Twitter (see section 5.1): it is possible to aggregate the tweets of one's connections within the service, to tweet one's LinkedIn status, and to update one's LinkedIn status from Twitter.

The culture on LinkedIn is geared towards personal network building and personal advancement rather than discussion. Even so, the career development angle suggests it has potential for promoting training-related activity. LinkedIn is heavily used by professional associations, alumni groups, and business networks – including some data management organisations, such as the Association of Clinical Data Management. It is also used in conference and event promotion, for example by Online Information. Users of the service complete a CV-like profile when they sign up; these profiles could be used to gauge the visibility of the DCC among its target communities.

Ning. Ning is a platform for creating small, themed social networks, offering a portal-like range of functionality that can be tailored to requirements.³⁰ These networks can be open or restricted to members only. The functions on offer include blogging, microblogging, chat, forums, image and video sharing. The microblogging function allows cross-posting to Twitter. Ning does not support aggregating content from other services, however.

29. LinkedIn Website, URL: <http://www.linkedin.com>|<http://www.linkedin.com/>

30. Ning Website, URL: <http://www.ning.com/>

Ning could potentially be useful, for example as a way of pooling DCC communications or reaching out to the Associates Network. Building and sustaining a Ning Network, though, requires similar effort to a Website, and the functions on offer overlap with those of Drupal, the platform used by the DCC Website.

There are other sites offering a very similar service to Ning, such as Elgg.³¹

Forums and bulletin boards. Forums and bulletin boards allow users to start a discussion topic within a particular category, whereupon other users post comments on that topic. The DCC has already experimented with this medium, using an installation of phpBB.³² On the one hand, forums are simple to administer, are widely implemented, and enable one to track how much each individual topic has been read. On the other hand, the registration process (in phpBB at least) is clunky, and the open range of topics means these tend to proliferate and be over-specific.

Forums work best where there is a strong group that monitors the forum closely, answering queries, and a smaller group that visits and asks questions. In the aforementioned experiment, the DCC failed to be the strong group that hangs out and answers well enough to make the DCC Forum somewhere that people naturally look. Currently the role is better served by alternatives such as the Research Data Management and Associates Network email lists.

Google Wave. Google Wave is a real-time collaboration platform, described by Paul (2010) as 'e-mail, instant messaging, an online collaboration tool and a wiki all rolled into one service'.³³ Further features may be added using plug-in widgets. In June 2010, Google released a bookmarklet and Chrome extension, WaveThis, which lets users launch a discussion in a collaboration space ('wave') directly from the site users want to share (Boulton, 2010b).

It is technology in a very early stage of development, and the publicity surrounding it has not succeeded in giving people a clear understanding of its purpose. It had attracted over 1 million active users by May 2010 (Boulton, 2010a) – not that many considering its high profile launch in 2009 – but remains an option for collaboration on resource development, between DCC sites or with partners.

Recommendations. Facebook is the most popular discussion and networking site in the world, but is used mainly for recreation rather than work purposes. LinkedIn is aimed squarely at the professional market, and while it has far fewer users than Facebook it is still relatively popular. It is therefore recommended that the DCC set up a corporate account on both of these platforms, and after a trial period decide whether it is worth maintaining a presence in both, or whether one should be retired.

31. Elgg Website, URL: <http://elgg.org/>

32. DCC Forum, URL: <http://forum.dcc.ac.uk/>

33. Google Wave Website, URL: <http://wave.google.com/>

5.4 Cross-posting, aggregation and monitoring

Cross-posting tools enable automatic posting of the same text to a range of different services, including blogs and microblogs. Monitoring tools track responses to that activity in one or more services. Aggregation tools track posts across a range of services, or feeds from them, and consolidate them in a single feed that may be embedded in a website.

5.4.1 Cross-posting tools

Ping.fm. Ping.fm is a free service enabling users to post to multiple other services such as blogs, microblogs and sites with news/status updates.³⁴ Among the supported services are Twitter, Facebook, LinkedIn, FriendFeed, WordPress, Blogger, Delicious and Flickr. The principal advantage of using it is that avoids duplication of effort in posting the same message to all these services individually. Of course, it is effectively a broadcasting tool – it does not deal with any ensuing conversation – but it could complement an aggregation service that does. Neither does it address the effort needed to ‘be sociable’ in the relevant networks, for example to build connections to other people.

Ping.fm is probably the best known cross-posting tool, but there are a handful of others, for example Gwibber, hellotxt and tarpipe.³⁵

5.4.2 Monitoring

Karthik (2010) provides a review of ten tools for social media monitoring. It is recommended that further assessment of the most suitable tools for the DCC to use be postponed until it is clear which services need to be monitored.

5.4.3 Aggregation

FriendFeed. FriendFeed is an aggregation service that makes it easy to share with ‘friends’ any content from a wide range of services, including other social media sites, blogs, microblogging and bookmarking sites, or RSS/Atom feeds; it is also possible to post original messages.³⁶ This content can easily be commented on by the other FriendFeed users who subscribe to it. The content presented to users is ranked in an effort to bring the most interesting content to the top: this is determined by how many of one’s friends indicate they ‘like’ it, how much they are commenting on it, how many people are commenting on it generally, and so on. FriendFeed is useful for filtering activity on a range of topics across a variety of services, combining these in a single feed, but does not have built-in facilities for pushing content to other services.

34. Ping.fm Website, URL: <http://www.ping.fm/>

35. Gwibber Website, URL: <http://gwibber.com/>; hellotxt Website, URL: <http://hellotxt.com/>; tarpipe Website, URL: <http://tarpipe.com/>

36. FriendFeed Website, URL: <http://friendfeed.com/>

It is possible to set up 'groups' within the service, but these act approximately as subject headers: anyone with an interest in the topic can subscribe to them, and there is little collaborative functionality associated with them. Nevertheless, there are active groups that may match some of the DCC's target communities: researchers (Open Science, eResearch, Science 2.0, Science Apps), data scientists and data centre staff (Big Data, Data Mining, Data Portability, Data Center, Data Visualization, Open Data, Hot Data, Linked Data). Participation seems to be reasonably strong in some life sciences, but still only a minority use the service.

FriendFeed was acquired by Facebook in 2009. Since then a founder of FriendFeed has become CTO of Facebook, and Facebook has begun to incorporate some of FriendFeed's functionality. What this means for FriendFeed's future as a separate entity remains to be seen. An open source version of the Web server software underlying FriendFeed has been released under the name Tornado,³⁷ and there are the beginnings of a movement to build a community-owned alternative. It is not clear how successful this could be. A similar but less functional service aimed at scientists was launched in February 2010 under the name Sciencefeed.³⁸

Disqus. Disqus is a commenting platform that, it is claimed, works with any blogging software; it is designed to support conversations and sort out spam.³⁹ Described as a 'comment management system', Disqus comes in two forms: Disqus Comments, a drop-in replacement commenting system and moderation tool for Websites, including those using Drupal, and Disqus Profile, a service for centrally managing one's commenting profile across multiple sites. The former service provides a comment rating ('Like') button, and displays the number of ratings and who has made them. The latter service allows comments posted in a blog or website to be syndicated across other services; it also searches across services (Twitter, Facebook, etc.) and allows the results to be displayed and commented on at one's own site. These two features together make it possible to synchronise Disqus Profile with FriendFeed.

There are some points of concern about using Disqus Comments on the DCC Website. In a sense it relinquishes control of comment data from the site to a third party. The impact on server performance must be considered, since the interface needs to be loaded on any page that provides comments. Lastly, Disqus operates both free and paid-for services, implying that the free functionality may be limited or variable.

Recommendations. The DCC should set up a corporate FriendFeed account to aggregate content from the other services recommended by this report.

37. Tornado Web server Website, URL: <http://www.tornadoweb.org/>

38. Sciencefeed Website, URL: <http://www.sciencefeed.com/>

39. Disqus Website, URL: <http://disqus.com/>

5.5 Collaborative writing

Collaborative writing is a relatively common way of composing documents, but in a social networking context it takes on a new twist: the authors composing the document may be complete strangers to one another. In this section, therefore, we consider services that would allow the DCC to contribute to documents written by others, or allow those in the DCC's target communities to contribute to DCC documents.⁴⁰

For individual documents, several online Office suites are available that replicate the desktop document authoring experience in a more collaborative context. Perhaps the best known of these is Google Docs, although alternatives such as Microsoft's forthcoming Office Web Apps and Adobe's Acrobat.com provide very similar functionality.⁴¹ It is possible to upload and convert existing documents to Google Docs format: Microsoft Office .doc, .docx, .ppt, .pps, .xls and .xlsx, OpenDocument .odt and .ods, StarOffice .sxw, rich text, plain text, HTML and CSV format files are supported. Equally, it is possible to export edited documents to any of these formats. Versioning is controlled much as in wikis, although the latest version of the service allows real-time collaborative editing, so that users can see each other's edits as they make them. As noted in section 2.1, these online applications are not as powerful or functional as desktop office applications, so for complex or heavily branded documents the lead author may need to post-process an exported version of the final draft before publishing it. The lack of pagination in editing mode can make it hard to navigate long documents, though this can be worked around by splitting the document into several files and using a parent folder for setting permissions. The system of folders could also be used as a way of creating a collaborative editing space, where anyone could start a new document and collaborators could see this and participate without further notification or invitation. The disadvantage of this might be that, once a critical number of documents have been added, navigating and locating documents would become rather less intuitive than in, say, a wiki system.

Drupal, the software used to manage the Phase 3 DCC Website, has the facility to allow visitors to give feedback on the site's content. On a very simplistic level, a vote up/down module can be used to gauge the more popular items on the site, and those that should be revised or withdrawn. The comment module allows visitors to give free-text responses to content, potentially inspiring a future revision. These methods fall short of collaborative writing proper, of course; given that the Website is the DCC's expression of itself on the Web, the DCC needs full control and responsibility for the content of the Website.

Having said that, there is no reason why the DCC could not set up a separate Web space that is manifestly a collaborative writing environment. This could be either a dedicated corner of the existing site, continuing to use Drupal for user management and page editing, or a separate wiki. In order to be a resource that people would want to read and edit, the space would need to have a clear objective, such as to provide a Curation Reference Guide as a companion to the Curation Reference Manual. The Psychology

40. The DCC already has three methods by which staff can collaborate internally on documents: a project wiki, Basecamp Writeboards and unpublished pages on the Phase 3 Website.

41. Google Docs, URL: <http://docs.google.com/>; Adobe Acrobat.com, URL: <http://www.acrobat.com/>

Wiki provides as example of such an endeavour.⁴² It is hosted by Wikia, a free wiki provider supported by advertisements; Wikia has links with Wikipedia and uses the same MediaWiki software. The professed aim of the Psychology Wiki is to provide a library of collaboratively written textbooks of sufficient quality that they are recognised as citable resources by national psychology societies. The methodology so far has been to create a knowledge structure based around the APA Thesaurus of Psychological Index Terms, divided into 17 'books', and to populate it material from Wikipedia, filling in the gaps with freshly written material. Several to-do lists have been set up to focus effort in this regard. The next stage in the workplan is to gather inspiration for further articles from psychology dictionaries. Once this is complete, a comprehensive set of references will be compiled for each article; these references will then be used to update the articles, at which point the wiki will switch to a phase of ongoing updates. Once the wiki has become established, the organisers hope to set up a formal peer review process for articles and to make editing rights contingent on professional qualifications. Outside of the textbook effort, the wiki also provides some professional services, listing upcoming conferences, calls for papers and so on.

The Shot of Jaq Website provides another model for collaborating on articles.⁴³ The Scratchpad section of the site is in essence a public wiki, but as well as the usual features it automatically generates a hashtag for each page, of the form #soj_ followed by a short string of random characters. Each page features a sidebar that pulls in messages from Twitter and Identi.ca that use that hashtag, allowing those viewing/editing the page to see discussions about it.

Without doubt the largest and most successful online collaborative writing project is Wikipedia.⁴⁴ In April 2010, it contained around 15 million articles in over 270 different languages, with English accounting for 3.3 million articles. It is operated by the Wikimedia Foundation and edited by a community of volunteers. While anyone may set up an account for editing Wikipedia, there are different levels of permissions through which an editor may progress, and different roles that editors may take on, such as guarding a set of pages from vandalism or checking new articles for quality.

Wikipedia has considerable impact, with its articles often ranking highly in search engine results and resurfacing on sites such as Answers.com. It strives towards the character of a professionally compiled encyclopædia, with strict policies on editorial neutrality, verifiability and replication of existing text, though the quirks of the editorial process can lead to unevenness of tone, detail and quality. As the articles are volunteered rather than commissioned, their coverage is particularly strong in areas of popular culture and comparatively weaker in more specialist or academic areas. Having said that, some areas are receiving particular attention from specialists, most notably in Chemistry⁴⁵ and Pharmacology.⁴⁶

42. Psychology Wiki, URL: <http://psychology.wikia.com/>

43. Shot of Jaq Scratchpad, URL: <http://shotofjaq.org/scratchpad/>

44. Wikipedia English Home Page, URL: <http://en.wikipedia.org/>

45. WikiProject Chemistry page, URL: http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Chemistry

46. WikiProject Pharmacology page, URL: http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Pharmacology

The software underlying Wikipedia is MediaWiki; editing therefore normally takes place through a standard wiki interface, although many actions may also be performed via the MediaWiki API.⁴⁷ The wiki software provides some tools that may be of interest for statistical purposes, such as page history and versioning, and lists of edits made by a particular editor.

One of the challenges faced by Wikipedia is that it appears to offer a medium for individuals and organisations to promote themselves at very little cost. The policies put in place to prevent this can also make it difficult for experts to write about projects and research in which they are closely involved. As noted by DigitalPreservationEurope (see section 2.1), the personalities of individual editors have a strong influence on the editorial process, and since most editors are volunteers with limited time, negotiations can be tricky.⁴⁸

For an organisation justifying the effort it spends working on Wikipedia, the impact that any particular edit has is hard to judge on at least two counts. First, the natural unit for considering impact would be the article, not the edit, so it is hard for any individual editor to claim the credit for any given example of impact. Second, the citation of Wikipedia articles is generally discouraged, not least by Wikipedia itself, so many uses of an article will go unrecognised.

Recommendations. While collaborative writing of the kind considered in this section opens up a number of exciting possibilities, such as collaboratively developing a guide-book to digital curation or allowing tool developers to update DCC training materials themselves, without evidence of community demand for such things there is a danger they may not gather the required momentum.

Editing Wikipedia would certainly be worthwhile in terms of the impact such work would have, but the culture of that project better suits a horizon-scanning role than a self-promotional one. In other words, DCC staff could usefully write reviews of third-party digital curation initiatives for Wikipedia, but should not seek to make substantial contributions concerning the DCC or its activities. If such reviews also become a feature of the DCC blog, it is recommended that different staff contribute to each review, in order to provide different perspectives and avoid charges of plagiarism or self-citation.

5.6 Social bookmarking

Social bookmarking is the practice of saving the URLs of Web resources in such a way as to allow others to view, comment on and rate them. While some social bookmarking services specialise in a particular type of resource, such as online news articles or blog posts, others are more eclectic and allow most forms of Web resource to be bookmarked; it is these latter services that we consider in this section.

47. MediaWiki API documentation page, as implemented in Wikipedia, URL: <http://en.wikipedia.org/w/api.php>

48. See, for example, the debate over the proposed deletion of the DBpedia article, URL: http://en.wikipedia.org/wiki/Wikipedia:Articles_for_deletion/DBpedia

The DCC is already engaging with social bookmarking, having added an AddThis widget to each page on the Website. AddThis and the similar ShareThis service each provide a widget containing buttons for submitting a Web resource to most social bookmarking, blogging and microblogging services. This makes it easy for visitors to the Website to share resources they find useful. There is also, however, a case for the DCC forming its own collection of bookmarks to share with its target communities.

Delicious. Delicious (formerly del.icio.us) is one of the best known social bookmarking sites.⁴⁹ It uses a relatively simple model: each link is associated with a list of users who have bookmarked it, and a list of tags that users have applied to it. The former provides a measure of the link's popularity, and the latter a folksonomic classification for retrieving bookmarks by topic. Unlike other services, Delicious does not collect user comments about a bookmark, but the front page pulls in messages from Twitter related to the featured bookmarks.

Delicious provides two ways of defining custom aggregations of bookmarks. One is to build a network: a group of users whose bookmarks are of interest, or in Delicious terminology, of which one is a fan; users in one's network may be subdivided into bundles if more than one aggregation of this kind would be useful. The other is to subscribe to a list of tags to which one is interested; again these tags may be grouped into bundles for finer thematic control. Delicious also allows one to recommend a bookmark to a user in one's network.

A sizeable suite of tools has grown up around Delicious. There are extensions and bookmarklets for browsers to simplify bookmarking sites, and more sophisticated tools for syncing browser bookmarks with Delicious bookmarks. Several widgets are provided for inclusion in Web pages: displaying a user's most recently added bookmarks or most recently used tags; displaying statistics about a user's bookmarks, network and fans; displaying the number of Delicious users who have bookmarked the current page; and displaying the tags that Delicious users have associated with the current page. Finally, Delicious provides an API for automated read/write access, and a newsfeed mechanism for simple access to the read-only aspects of the API. The API is comprehensive in the functions it provides, and libraries are available for using the API from PHP, Java, Python, Ruby, Perl, LISP, and C#, as well as for backing up Delicious bookmarks to a MySQL database.

Digg and Reddit. In comparison with Delicious, the service provided by Digg and Reddit is focused more on recommendation and less on bookmarking for personal reference.⁵⁰ Users submit links to the service with a custom title and description. Other users can then vote for or against it, and can attach their own comments; the comments may themselves be voted up or down as well. Both Digg and Reddit include advertisement links, but these too are ranked by user votes.

Again, it is possible to build a set of friends (whose submissions one is interested in) and fans (interested in one's own submissions). Reddit uses an internal mechanism to share

49. Delicious Website, URL: <http://delicious.com/>

50. Digg Website, URL: <http://digg.com/>; Reddit Website, URL: <http://www.reddit.com/>

links with friends, while Digg provides shortcuts for sharing stories via e-mail, Facebook or Twitter. It is also possible to associate submissions with a particular topic. In Digg, the topics are a fixed list, whereas in Reddit, users may set up their own topics ('subreddits') and exert some control over how that topic is displayed on the Reddit site, who may associate submissions with the topic, and so on.

Both services have a small ecosystem of tools. Digg has its own URL shortening service for easily quoting submissions in other fora, and integrates well with Facebook. A Firefox extension is available for supplementing Google, Yahoo! and Bing search results with those from Digg. Meanwhile for Reddit, there are bookmarklets for submitting sites, visiting a random submitted site, and injecting a Reddit toolbar at the top of ordinary Web pages; a Firefox extension and an iPhone application are also available giving much the same functionality. Widgets are available for inviting users to submit the current page to Reddit, displaying the page's current Reddit score, and for displaying submitted pages (optionally filtered by user activity or topic). The source code for Reddit is also available. Both Digg and Reddit have comprehensive APIs allowing alternative interfaces to be used.

StumbleUpon. Unlike the services above, StumbleUpon does not have a front page displaying recent or popular links; rather, it tries to provide the reader with interesting pages one at a time, similar to the 'go to random page' functions of other services.⁵¹ The recommendation engine uses a combination of a user's chosen topics of interest, the pages in which a user has expressed an interest (either from those previously recommended or from normal browsing) and the user's age/sex/location, alongside those of others selected by the user (analogous to the 'friends' of other services). StumbleUpon does not rely on users classifying resources: it uses an automatic classifier for that. Advertisers can pay to have pages added to the recommendation list, but these pages are affected by user votes as well.

StumbleUpon does not have a public API, so it does not have a large number of supporting tools; there are however StumbleUpon extensions for popular browsers. Besides those, there are widgets for submitting the current page to StumbleUpon and displaying the number of times people have viewed the page through the service, and a separate URL shortening service with its own API and WordPress plugin.⁵²

Recommendations. The DCC currently has two mechanisms for alerting members of its communities to relevant resources. For more ephemeral items like news stories, the DCC has its own newsfeed, or the blog if a more reflective approach is required. For more persistent items, there is the external resources page.⁵³ The DCC newsfeed can be filtered and repurposed already, so the principal benefits of using a social bookmarking platform would be (a) to enhance the utility of the external resources page, and (b) to provide additional mechanisms allowing the DCC to discover new resources and news

51. StumbleUpon Website, URL: <http://www.stumbleupon.com/>

52. StumbleUpon URL shortening service, URL: <http://su.pr/>

53. DCC Website, 'Digital Curation Resources from outside the DCC', URL: <http://www.dcc.ac.uk/resources/external>

items, as a direct or indirect result of members of the DCC communities submitting them.

On the first point, in order to prevent duplication of effort it would be sensible either to export items on that page to the social bookmarking platform, or generate the contents of that page dynamically from the data on the social bookmarking platform. The former approach is more stable (in terms of what would happen if the service disappeared) but technically more difficult and removed from the social aspects of the exercise. The utility of the latter approach largely depends on the kinds of data that can be retrieved about each bookmark: comments and ratings contributed by the community would improve the current page, and if these can be gathered from a social bookmarking platform it saves having to drive traffic to the DCC Website first and deal with user accounts. This, and the second point above, indicate the need to choose a platform on which DCC communities are already a strong presence. Ironically, though, our survey results (section 3.3) show the only dedicated social bookmarking platform used to any degree by our communities is Delicious, which does not support ratings or comments though it does provide a form of popularity measurement.

With these issues in mind, it is hard to give an outright recommendation, but it would seem that as Delicious has the greatest impact among our communities and is well supported by tools, it would be worth investigating how it could be used to improve information exchange. Given that Delicious integrates well with other services, it is possible that including an additional platform in the DCC's social networking strategy may overcome the tension described above.

5.7 Sharing references to literature

While social bookmarking can be used by academics and researchers to keep abreast of developments in their discipline, they have specific needs that are not addressed by general purpose tools. Academic discourse already has a system in place to alert readers to related resources, so social bookmarking systems have evolved that are adapted to integrate with this system of citations and bibliographies.

While the distinctions are becoming increasingly blurred, three of the platforms in this section – CiteULike, Connotea and BibSonomy – would be best described as Web services with potential for desktop integration, while the other two – Mendeley and Zotero – are better thought of as desktop applications that integrate with a Web service.

CiteULike. CiteULike is a service provided by a small (UK) Bristol-based company with sponsorship from Springer.⁵⁴ References may be added to a user's account using a browser bookmarklet, using 'copy' links when viewing references added to the system by others, or by filling out a submission form. Bulk import is also possible from BibTeX or RIS files. Metadata can be automatically scraped from the pages of 85 different publishers and repositories, and in many cases the URLs of the resources can be cleaned

54. CiteULike Website, URL: <http://www.citeulike.org/>

to improve stability. PDFs can be uploaded or copied into the system via URL. Users can apply personal or public tags to references, and add ratings and reviews. Tags are used to delimit batch operations as well as to classify resources.

Every user is given a blog to which references can be pushed. CiteULike supports groups, which can be either public or private. Each group has its own blog and forum, and users can share references with the group or with the entire site; PDFs can be shared within private groups but not public ones. In addition, the service supports direct connections between users, allowing them to send messages to each other through the system and to mark recommendations for each other using special 'for:' tags. It is possible to track references added by specific users, or under specific tags, or written by specific authors.

Various types of searches are available: popular references from the last seven days, latest additions, users with a specific research interest, users with a similar set of references, etc. Tables of contents for many journals are available to browse. It is possible to use CiteULike in conjunction with OpenURL and EZProxy. As a spam prevention measure, references and metadata entered manually are treated as private and therefore not shared with other users.

References may be exported to BibTeX, RIS, JSON, PDF, RTF and plain text. As well as user blogs, they may also be pushed to Twitter and Delicious. The service does not expose an API, but it is possible to synchronise references with a Mendeley account, and the application SyncUThink may be used to synchronise CiteULike references with a local PDF collection. Various RSS feeds are available.

Connotea. Connotea is a service provided by Nature Publishing Group.⁵⁵ References may be added to a user's account using a browser bookmarklet, using 'copy' links when viewing references added to the system by others, or by filling out a submission form. Bulk import is also possible from BibTeX, RIS, EndNote XML, MODS, ISI Web of Knowledge and Firefox bookmark files. Metadata can be automatically scraped from the pages of 16 different publishers and repositories. Locally saved PDFs (etc.) may be uploaded. Users can apply tags and comments to references, and attach notes to tags explaining their meaning/usage.

Each user and group is given a page on the Connotea Community Pages wiki. Connotea supports both public and private groups. References may be shared within a group or with the entire site. It is possible to track references added by specific users or under specific tags. Searches are available for popular references, latest additions, popular tags, related tags, users with a similar set of references, etc. It is possible to use Connotea in conjunction with OpenURL, and to login using OpenID. A spam prevention mechanism exists for enforcing some references as private, but how this operates this is unclear.

References may be exported to BibTeX, RIS, EndNote XML, MODS and RDF. An API has been published for interacting with the service, and libraries are available for Perl, Python and Ruby. As well as the bookmarklet for adding references, others exist for

55. Connotea Website, URL: <http://www.connotea.org/>

retrieving and adding comments about a resource currently viewed in the browser. Many types of data are syndicated using RSS.

BibSonomy. BibSonomy is a service provided by the Knowledge and Data Engineering Group of the University of Kassel.⁵⁶ References may be added to a user's account using a browser bookmarklet, using 'copy' links when viewing references add to the system by others, or by filling out a submission form. Bulk import is also possible from BibTeX and Firefox bookmark files, and from Delicious. Metadata can be automatically scraped from the pages of 75 different publishers and repository systems, and parsed from the Handle records of resources with DOIs. Locally saved PDFs (etc.) may be uploaded. Users may apply tags and comments to references, and specify hierarchical relationships (narrower, related, broader) between tags. A 'basket' is provided for delimiting batch operations.

Each user has their own group which they populate with 'friends'. BibSonomy administrators can set up other groups on request. References may be shared within groups or the entire site; PDFs may only be shared within groups. It is possible for users to mark recommendations for friends using special 'send:' tags; references may also be highlighted for an entire group. It is possible to track references added by specific users, or under specific tags, or written by specific authors.

Various types of searches are available: latest additions, users with a similar set of references, etc. It is possible to sort references by date or calculated relevance (using a 'FolkRank' algorithm). It is also possible to browse related tags (ones that often occur with the tag), similar tags (ones used in a similar context), and related concepts (in terms of the manually specified relations). It is possible to use BibSonomy in conjunction with OpenURL, and to login using OpenID.

References may be exported to BibTeX, EndNote, RDF, HTML, JSON, and Firefox bookmark files. They can also be pushed to Delicious. An API has been published for interacting with the service, with libraries available for Java and Python, and so far BibSonomy has been integrated with JabRef, Kibitex, Gnome Deskbar and CiteSmart. Code has been published to allow references to be pushed to BibSonomy from library OPACS and pulled from BibSonomy into WordPress, Confluence, Typo3 and Moodle; a workaround exists for synchronising with Zotero. A JavaScript library is available for using BibSonomy data to create tag clouds on Webpages. A Firefox addon allows Google search results to be supplemented by BibSonomy searches and submission links. Various RSS feeds are available.

Mendeley. Mendeley is an application and Web service developed by a small London-based company supported by former senior figures from Last.fm, Warner Music Group and Skype.⁵⁷ References may be added to a user's collection using a browser bookmarklet or through the desktop application, either by pointing it at a saved PDF, looking up a document using CrossRef, PubMed, ArXiv or Google Scholar, or by manual entry. Bulk

56. BibSonomy Website, URL: <http://www.bibsonomy.org/>

57. Mendeley, URL: <http://www.mendeley.com/>

import is possible from BibTeX, RIS and Endnote XML files. Metadata can be scraped from Web pages and PDF files, or looked up from 50 bibliographic databases. Up to 500 MB of PDFs may be synchronised with the Web service (though premium accounts will have larger quotas). Users can apply tags, comments and annotations to references, mark favourite references, and indicate whether they have been read. References may also be grouped into collections independently of tags.

Each user is given a CV page they can populate with their own publications. Mendeley supports private groups of up to 10 users (larger groups will be supported by premium accounts). It is possible to share references, PDFs and annotations within groups, as well as assign tasks and hold discussions. It is also possible to track the references added by specific users.

Various forms of search are possible: popular authors (generally or in a given field), popular references, popular topics, popular journals, related references, etc. Since the application has an integrated PDF viewer it is also possible to retrieve statistics about how much a given paper or author has been read.

References may be exported to BibTeX, RIS, EndNote XML and rich text. It is possible to synchronise Mendeley with CiteULike and Zotero, and to push references to Word and OpenOffice. The desktop application can be used to manage locally stored PDFs, including indexing and full text searches (using OCR to interpret scan-image PDFs). Various RSS feeds are available from the Web service.

Zotero. Zotero is a Firefox addon developed by the Center for History and New Media, George Mason University with funding from the Andrew W. Mellon Foundation, the Institute of Museum and Library Services, and the Alfred P. Sloan Foundation.⁵⁸ References may be added to a user's collection via a button in the Firefox address bar or manually through a form interface. Bulk import is possible from BibTeX, RIS, Bibix, MODS or Dublin Core RDF files. Metadata can be scraped from Web pages and PDF files, or retrieved from bibliographic databases. Users can apply tags and comments to references; referenced Web pages may also be annotated. References may be grouped into collections independently of tags.

Each user is given a CV page they can populate with their own publications. Zotero supports hidden, private and public groups; references, files and notes may be shared within groups, and group collections may remain private to the group or exposed publicly. It is also possible to track the references added by specific users.

References may be exported to BibTeX, RIS and rich text. It is possible to synchronise Zotero with Mendeley, and to push references to Word and OpenOffice. It can even replace special markup in RTF documents with fully formatted citations and bibliographies. It can index and perform full text searches on locally stored PDFs, provided some additional dependencies are installed.

Recommendations. If the summaries above show anything, it is that there is little to choose between the various platforms that are on offer, due no doubt to a healthy

58. Zotero, URL: <http://www.zotero.org/>

cross-fertilisation of ideas. Of the three mainly-Web services, Connotea is probably the weakest in terms of total feature set. CiteULike has the advantage of integration with Twitter and Mendeley (and therefore might be used to circumvent the latter's restriction on group size), but as it is biased towards peer-reviewed, mainstream journal publications it may be less suitable for the digital curation discipline, in which grey literature is particularly prominent.

Our survey (section 3.4) showed that the only specialist tool among these to gain any traction is Mendeley. Given its impressive feature set and ability to synchronise with other tools, there is a lot to recommend it. The principal drawback is the 10-user limit to group size, which is too small to be useful for the DCC as a set of individual staff, and too small to be used in conjunction with a corporate account for outreach into the community; the only realistic use would be in the context of colleagues collaborating on a document. As already mentioned, a possible workaround for this would be to use the group facility of a supported alternative system such as CiteULike or Zotero.

5.8 Sharing images

Flickr. Flickr is probably the best-known photo sharing site.⁵⁹ It is free to use in basic mode, but requires a subscription for advanced features. It is owned by Yahoo, and a Yahoo ID is needed to sign in, which is a touch inconvenient, but unsurprising given that lots of Google tools require a Google account.

Lots of APIs are available via the Flickr App Garden.⁶⁰ Users can create 'groups' related to a particular theme.

Flickr Commons has shown that adding images to Flickr from museums and other sources can lead to valuable additional metadata (and even other images) being contributed by the public. On the other hand, there is little evidence of this activity in the sciences.

Photobucket. Photobucket claims to be 'the most popular site on the internet for uploading, sharing, linking and finding photos, videos and graphics'.⁶¹ A comprehensive API is available.

ImageShack. ImageShack can be used either with an account or anonymously.⁶² An API does exist, but only privately: developers must request access. A toolbar is available for IE and Firefox, making it easier to upload pictures and view pictures already uploaded.

59. Flickr Website, URL: <http://www.flickr.com/>

60. Flickr App Garden Web pages, URL: <http://www.flickr.com/services/api/>

61. Photobucket Website, URL: <http://photobucket.com/>

62. ImageShack Website, URL: <http://www.imageshack.us/>

DeviantART. DeviantART describes itself as ‘where art meets application’.⁶³ It is aimed at the computer-literate artist, and so hosts not only photographs but flash animations, application skins, desktop themes and so on. While the name may suggest the service hosts more outré content than it actually does, it still may not be an ideal fit for the DCC.

Fotolog. Fotolog has both a free and subscription memberships.⁶⁴ Free members can only upload one image per day, while Gold Camera members can upload up to six per day. Fotolog does not have a public API, though it may have a private one. As with Flickr, it is possible for users to create ‘groups’ related to a particular theme.

Recommendations. The DCC does not create many images, but it has produced some important graphics (e.g. the Curation Lifecycle Model) and collects photographs from important events such as the International Digital Curation Conference (IDCC). Of the sites reviewed, Flickr would seem to have the best combination of reputation, user base, functionality and reach. It is therefore recommended that the DCC set up a corporate Flickr account, and upload to it promotional photography, diagrams, and so on.

5.9 Sharing presentations

SlideShare. SlideShare is probably the best-known site for sharing presentations.⁶⁵ It allows one to share presentations publicly and privately. Slides may be uploaded as PowerPoint, PDF or Word. Presentations are displayed using Flash and may be embedded in other Web pages; a text-only version is also generated. An audio track may be added. Others can comment on the presentation. On the one hand, it makes it very easy to display slides on Web sites, in blogs, etc., and it is possible to use the online version as a backup if there are problems displaying presentations live at the venue. On the other hand, having the same presentation available from both the DCC Website and SlideShare may adversely affect its visibility in search engine rankings.

SlideShare has a comprehensive open API. The site makes it easy to share slides on Twitter, Facebook, Google Buzz, WordPress, Blogger, etc. Page views are counted, which can help with recording impact. Pages can be ‘liked’ on Facebook.

Myplnick. The stated goal of Myplnick is to host presentations that are as lively as ‘flicks’.⁶⁶ In practice it seems to appeal to the more amateur end of the market. The site displays intrusive advertising and spam is rife in both uploaded presentations and comments. On the positive side, buttons are provided for sharing presentations on

63. DeviantART Website, URL: <http://www.deviantart.com/>

64. Fotolog Website, URL: <http://www.fotolog.com/>

65. SlideShare Website, URL: <http://www.slideshare.net/>

66. Myplnick Website, URL: <http://www.myplnick.com/>

Facebook and Twitter, and the site shows statistics such as the number of views, number of downloads, average number of slides seen in each viewing, and the average duration of each viewing.

SlideBoom. SlideBoom is a showcase for iSpring PowerPoint conversion software, and is aimed at the corporate market.⁶⁷ With free accounts, only presentations up to 100 MB can be uploaded. Unlike SlideShare, SlideBoom can reproduce PowerPoint animations, slide transitions and other special effects in the Flash version; this comes at the expense of flexibility as the site only accepts PowerPoint files. The display of the text transcript and slide notes is a little neater than on SlideShare. Code is provided for embedding SlideShare presentations in third party Websites, but for WordPress blogs a special plugin is provided.

The site displays statistics on the number of times a presentation has been viewed, downloaded and embedded, while a Pro account promises rather more detailed statistics such as activity over time and geographical locations of viewers. The blog for the site has not been updated since July 2009, suggesting development may have halted.

SlideServe. SlideServe is a showcase for DigitalOfficePro's PowerPoint conversion software.⁶⁸ The interface could be described as basic or uncluttered, depending on one's perspective. As with SlideBoom, SlideServe can reproduce PowerPoint animations, slide transitions and other special effects in the Flash version, at the expense of supporting other file types. In terms of site functionality, it does not appear to offer anything that SlideShare does not do better, apart from the facility to upload presentations to YouTube. Uploads are limited to 50 MB; it does not provide a download facility.

Code is provided for embedding SlideServe presentations in third party websites. Statistics are shown for the number of times a presentation has been viewed and how it has been rated; three other measures are also provided, but the icons are not explained.

Recommendations. In terms of reputation, user base, functionality and site design, SlideShare is the best fit for the DCC. It is recommended that individuals set up their own SlideShare accounts, and the DCC set up a special Group for DCC staff and their (DCC) presentations.

5.10 Sharing video

YouTube. YouTube was acquired by Google in 2006 and remains the most visited video sharing site; indeed, it is the world's third most visited Website after Google and Facebook.⁶⁹ Organisations can create their own 'channels', thereby controlling the way their content is presented. Resolutions from 240 lines to 1080 lines are supported.

67. SlideBoom Website, URL: <http://www.slideboom.com/>

68. SlideServe Website, URL: <http://www.slideserve.com/>

69. YouTube Website, URL: <http://www.youtube.com/>

Videos uploaded to YouTube can easily be embedded in other webpages using cut and paste code. There are several player APIs and a Data API that allows access to most of the functions of the Website.⁷⁰ Page views are counted, allowing the impact of videos to be gauged. Comments may be posted on each video.

Dailymotion. Dailymotion is a video sharing site based in France.⁷¹ It supports resolutions from 240 lines to 720 lines and video durations up to 20 minutes, although MotionMakers (creators of original content) and Official Users (record labels, film studios, etc.) are not restricted in this way. Registering as a MotionMaker does not appear to have cost implications. There are open APIs for embedding Dailymotion videos in third party Websites. Page views are counted, allowing the impact of videos to be gauged. Comments may be posted on each video.

Metacafe. Metacafe describes itself as ‘online video entertainment’.⁷² All content is reviewed by volunteers before going live to ensure that it has a suitable age ‘certificate’, that it meets the site terms and conditions, and that it is not a duplicate. Video metadata (titles, tags, descriptions) may be edited by all registered users.

Metacafe provides RSS feeds of videos with a certain tag or uploaded by a particular user, as well as an API for performing more complex queries and retrieving the results as RSS. Generic code is available for embedding Metacafe videos in third party Websites, but tailored code is also available for specific platforms (e.g. Facebook, WordPress, Blogger). Page views are counted, allowing the impact of videos to be gauged. Comments may be posted on each video.

Veoh. Veoh went bankrupt in February 2010, but has since been bought by Qlipso, an Israeli startup company.⁷³ It hosts both user-generated content and TV and movie content. An private API is available only to users who have registered with the site and completed a developer profile. Page views are counted, allowing the impact of videos to be gauged. Comments may be posted on each video. Veoh has a history of blocking access from certain countries, and hence may not be an ideal home for DCC content.

Crackle. Crackle, formerly Grouper, is owned by Sony and positions itself as both a showcase for new talent and a platform for TV and movie content.⁷⁴ Given that the emphasis is on entertainment, rather than information or education, Crackle does not provide a natural home for DCC content. Furthermore, it has a chequered history of blocking access from countries outside the US, making its reliability for a UK-based organisation questionable.

70. For examples of sites that make use of these APIs, see the YouTube API Case Studies Web page, URL: <http://code.google.com/apis/youtube/casestudies/>

71. Dailymotion Website, URL: <http://www.dailymotion.com/>

72. Metacafe Website, URL: <http://www.metacafe.com/>

73. Veoh Website, URL: <http://www.veoh.com/>

74. Crackle Website, URL: <http://www.crackle.com/>

Vimeo. Vimeo exclusively hosts user-generated content.⁷⁵ The site is arguably less brash and advert-orientated than that of YouTube. It supports resolutions from 240 lines to 720 lines. The site has both free and paid-for (Plus) membership options, with the latter offering more generous upload limits, higher quality transcoding and a few other features.

The site offers both a simple, open API and an advanced API for registered applications only; it also provides code for embedding videos in third party Websites, and documentation for controlling the Vimeo video player with JavaScript and embedding it in Flash applications. There is a forum for peer support on the APIs.⁷⁶

Page views and 'likes' are counted, allowing the impact of videos to be gauged; with free accounts, it is possible to see how the views and 'likes' were distributed over the past seven days. More detailed statistics are available to Plus members. Comments may be posted on each video; this ties in with the social networking functionality (contacts, 'likes', forums, groups, etc.).

Recommendations. Of the sites reviewed, Vimeo is probably the best fit for the DCC in terms of tone and functionality, although it does not have the reach of YouTube. It is therefore recommended that the DCC set up corporate accounts on both these services and trial them in parallel. While the DCC does not currently produce much video output, it has the potential to do so, particularly with respect to the Research Data Management Forum, training courses and other events. The DCC should therefore invest in a small HD video camera for this purpose.

75. Vimeo Website, URL: <http://vimeo.com/>

76. Vimeo API Forum, URL: <http://vimeo.com/forum:API>

6 Proposed social networking strategy

Building the curation community and stronger connections

The DCC website is evolving in line with the change of emphasis in DCC Phase 3 compared with Phase 2, on building curation capabilities by leveraging and contributing to the activities of our target communities, aiming to add structure and coherence – rather than trying to serve as the font of all knowledge and expertise on curation for all research communities and service providers. The DCC website in Phase 2 served as a ‘library of good practice’ in research data management and curation, and in Phase 3 has more emphasis on becoming a ‘hub for exchanging, filtering, and amplifying good practice’. That ties in with broader objectives and requires an explicit strategy for using social networking capabilities effectively. An important factor here is that we need to strengthen our connections with a more focused cross-section of communities or target audiences, but have fewer people to do that and no dedicated Webmaster to coordinate online activity. The social networking strategy should comprise the following.

1. Target communities
2. Aims, outcomes and measurable objectives
3. Communication patterns
4. Recommended platforms and tools
5. Intellectual property policy
6. Operating processes
7. Implementation Plan

The strategy presented here is a working draft, as its first two steps are highly interconnected with the ongoing review of the Phase 3 Business Plan, Communication Strategy and DCC Website.

6.1 Target communities

Drawing on the consultation survey with DCC directors and staff, we can identify the following target communities along with their primary interests.

- *Data creators and data scientists*: researchers at all career levels with an interest in sharing research data or methods online, typically in interdisciplinary data-intensive fields.
- *Research Data Managers/Curators*: typically working at research group or department level, or providing subject-based services operated by funding-body supported data centres.
- *Library/IT Service Support Staff*: (including ‘data librarians’) responsible for delivering national or HEI-based data support services and infrastructure, including HEI repositories and funding-body supported data centres.
- *Research Administrators and Support Staff*: in HEI services responsible for research policy, codes of practice, research assessment and impact, PR and communications, managing research information systems.
- *Funding and Support Agency Staff*: responsible for developing policy and systems to support HEIs’ research infrastructure
- *Consultants*: private sector actors supplying data management/curation consultancy, products or services to any of the above.

6.2 Aims, outcomes and measurable objectives

The measures taken to implement social networking should address aims, outcomes and objectives. These are drafted below.

Aims	Outcomes	Objectives
Inform, advocate, educate	Greater awareness of what we do and the events, resources and training we provide.	1. Increase the number of <ul style="list-style-type: none"> • visitors to the DCC website • resource downloads • registered event participants
	Better support for DCC staff horizon scanning role.	2. Improve staff ability to effectively identify relevant issues, external events and resources.
Motivate	Improved satisfaction with DCC online activities	3. Improve the rating of DCC’s online presence by our target communities
		4. Improve DCC staff satisfaction with online role.
Engage	Associates Network membership more inclusive of our target communities	5. Establish presence on selected social networking platforms, linking to DCC online activities.
		6. Increase the number of Associates Network members from each of our target audiences.

	More online exchanges between DCC staff, community activists and champions	7. Engage more frequently in external online discussion.
	Associates Network contributing more expertise as 'Shadow DCC'	8. Provide Associates Network with capabilities to contribute to DCC activities, through comments and ratings of resources, and advice to the broader community.
Reward	Associates Network obtaining more benefits and increased satisfaction from membership	9. Raise the profile of contributing Associates Network members' curation activities and interests.

The objectives may be measured as follows. The cycle on which such measurements will be made needs to be determined.

Objectives	Measures
1. Increase the number of <ul style="list-style-type: none"> visitors to the DCC website resource downloads registered event participants 	<ul style="list-style-type: none"> Visits as measured by Google Analytics Downloads as measured by Google Analytics Event participants identifying DCC site or social networking presence as their source of event information.
2. Improve staff ability to effectively identify relevant issues, external events and resources.	<ul style="list-style-type: none"> Examples of identified issues, events, resources
3. Improve the rating of DCC's online presence by our target communities.	<ul style="list-style-type: none"> Website user satisfaction survey responses Number of tagged/bookmarked resources by relevant groups or social networks
4. Improve DCC staff satisfaction with online role.	<ul style="list-style-type: none"> Staff satisfaction survey responses
5. Establish presence on selected social networking platforms, linking to DCC online activities.	<ul style="list-style-type: none"> Number of contacts (fans, friends/followers) made through social networks.
6. Engage more frequently in external online discussion.	<ul style="list-style-type: none"> Number of posts made by DCC staff on social networks/blogs
7. Increase the number of Associates Network members from each of our target audiences.	<ul style="list-style-type: none"> Membership numbers and demographics

<p>8. Provide Associates Network with capabilities to contribute to DCC activities, through comments and ratings of resources, and advice to the broader community.</p>	<ul style="list-style-type: none"> • Number of active contributors (~ 1%) • Number of contributions to resources • Number of comments and ratings of resources
<p>9. Raise the profile of contributing Associates Network members' curation activities and interests.</p>	<ul style="list-style-type: none"> • Member interviews, profiles are published on DCC website • Contributors' other relevant resources are highlighted through DCC website or social network presence.

6.3 Communication patterns

Figure 6.1 visualises the DCC's interaction with its target communities in terms of the kinds of message we aim to communicate, the responses we need in order to meet our objectives and obtain desirable outcomes, and the main actions needed to accomplish this.

Meeting the aims depends on how we convey the DCC's 'message', which should inform and educate our target communities, advocate the actions we believe are needed. Interaction should engage their attention and participation, reward contributions, and motivate further interaction. Four steps are identified:

1. Establish a presence on selected platforms, as outposts of the DCC website.
2. Use the capabilities these platforms afford to listen and connect with others.
3. Amplify the desirable developments and repost these to our selected outposts.
4. Filter and aggregate relevant content we find through exchanges with the community.

6.4 Recommended platforms and tools

The selection of social networking platforms and tools reflects the four steps above, considering also

- our assumptions about the role of social networking/media in DCC activity;
- current DCC website capabilities – how the community can respond to our activity, and how we currently respond to activity in the community using the site;
- current social networking tools we are using, and the ease of integrating new tools into existing working practices.

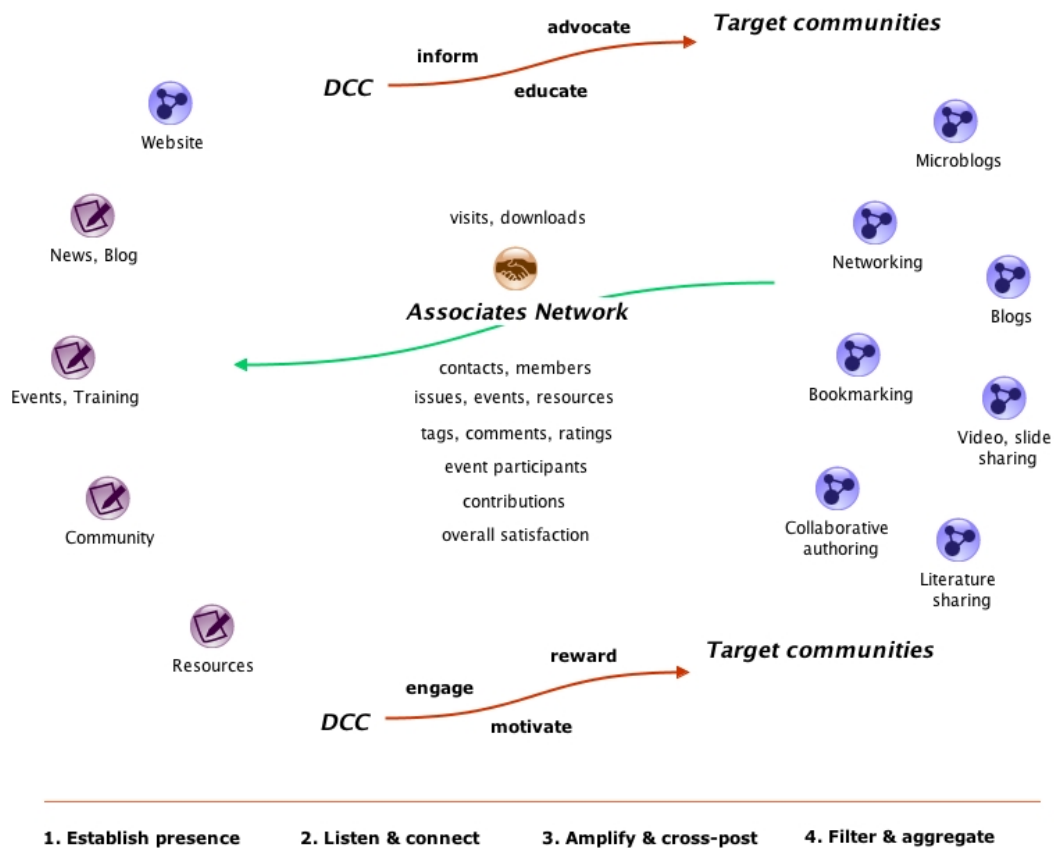


Figure 6.1: Communication patterns

6.4.1 Assumptions

1. The DCC Website will continue to be organised much as it currently is, but its content will increasingly be drawn from external sources ('filtered'), with selected items highlighted and commented on ('amplified').
2. We should periodically review the presence we need on current social networks relevant to our target communities, rather than picking one we believe to be a 'winner'.
3. Duplication of effort should be avoided.
4. To support our 'horizon scanning' role, we can aggregate from selected networks onto the Community section of our site, and amplify relevant posts by reposting in our own news items, comments and blogs.
5. Provided that external posts include trackback links back to our Website, this reposting should encourage traffic to it, providing higher visibility and site ranking.
6. We will individually have our own preferences for social networks, blogs, and so on we find interesting/relevant for horizon scanning, and should share these

preferences internally. An editorial group could review the selection of sources to aggregate periodically.

6.4.2 How the community can respond to the DCC website

The following table summarises capabilities currently implemented in the DCC website.

DCC website content we initiate	How community response afforded	Impact of response on DCC site	Issues – then and now
News: post news item	Comment facility in old site, also available in new site but restricted to news/blog items	Comments viewable	Comments rarely if ever used – possibly a trust issue – responding so publicly carries risks but little reward? Guidance needed on how to respond to criticism? Additional rating or ‘like’ function for users?
	Tag/bookmark article	None visible, but potentially higher traffic	Show number of saves to each service, e.g. how often tagged in Twitter, Delicious? (See Guardian Website)
	Subscribe to newsfeed	None visible, but potentially higher traffic	Well-known lists such as CNI take RSS feed – should we not measure usage?
News: post blog item	Comment on blog item	In old site news articles highlighted selected blog posts. In new site news and comments are integrated.	Blog has been highly rated according to Technorati ranking, etc. but needs sustained effort.
Event: Announcement/call for participation	Register attendance	None directly – handled separately by UKOLN	Show attendance numbers automatically or continue to do manually in, e.g. news items reporting on IDCC?
	Submit article or presentation	None directly – handled separately by UKOLN	Show number of submissions? Has also been done manually for IDCC.

	Provide evaluation/ feedback	None directly – handled separately by UKOLN	Show results? Who would monitor and respond to criticism?
	Provide record of presentation	Downloadable presentations	Show number of downloads?
Resource: publish report	Comment on page	Visible in principle	Never taken up on old site
	Review external resource	Visible in principle	Never taken up on old site
About: invite contact	Invite to join Associates' Network	None	Show current number of members and something about who they are? (consent issue here)
	Email addresses, helpdesk form	None	Link to FAQs? Collect 'was this answer helpful?' responses?

6.4.3 How we currently respond to external online activity

The following table summarises the website capabilities for DCC staff to report on activities of our target communities we find elsewhere on the web, or email lists.

External online activity	How DCC response afforded	Impact on DCC site	Issues
Post on external website about their event or resource	Phase 2: Submit a news article via web master. Phase 3: Post a news article or blog item.	External news distinguished from DCC news. Blog previously separate, now integrated	Previously depended on Webmaster to take initiative to filter/amplify – what is our approach now? Blogger site highly rated, feeds into Nature Networks et al. – how to integrate?
Post on external website about DCC event or resource	Reciprocate by linking to their site	Visible in useful links (Phase 2), Associated projects (Phase 3)	Are we proactive enough, e.g. in monitoring who has mentioned us?
Make tool or resource available	List as external tool/resource	Visible in resources section	Previously depended on Webmaster to take initiative to filter/amplify – what is our approach now?

DCC Forum post (Phase 2)	Respond within forum as individual user or moderator	Posts and comments visible	Low usage, effort to register, not well integrated with main Web presence
DCC Associates list post	Respond by email as list member	None	Web access provided by third party.
Anything else	Submit a news article via Web master	Visible as external news item or resource/tool	Previously depended on Webmaster to take initiative to filter/amplify – what is our approach now?

6.4.4 Recommendations: a hub for filtering and amplifying good practice

The report has reviewed and recommended a selection of platforms or tools in the following categories, based on their fit to objectives, capabilities and take-up in target audiences.

Microblogging *Twitter* account @dcc to publicise resources and events. *BirdHerd* to enable individual DCC tweeters to each post to it.

Blogging *DCC Website* for blog posts, mirrored on *Blogspot* account. Repost and comment on topical blog and forum posts on, say, *Nature Network* when relevant, e.g. after search on *Google Blogsearch* and *Technorati*.

Discussion and networking *Facebook* and/or *LinkedIn* to build community around events and training.

Cross-posting, aggregation and monitoring *FriendFeed* to aggregate from and monitor other services, e.g. blog RSS/Atom feeds, *Twitter*. To cross-post to other services use *Ping.fm* or similar.

Collaborative writing *DCC wiki* to collaborate on draft reference materials and their peer review. Consider contributing reference materials to *Wikipedia* using an individual account; monitor/record contributions using RSS/Atom feed.¹

Social bookmarking Provide *DCC Website* users with the ability to share items on *Delicious*; use *Delicious* for horizon scanning. Monitor/record contributions using RSS or JSON feeds,² and use tag 'data_curation_resources'.

Sharing references to literature For horizon scanning consider searches on *CiteU-Like*, *Zotero* and *Mendeley*. Synchronise references held in such services where possible. Use individual accounts and ensure IJDC papers/articles are added as they are read.

1. Example *Wikipedia* contribution feed: <http://en.wikipedia.org/w/index.php?title=Special:Contributions/username&feed=atom&limit=50>

2. Example *Delicious* contribution feed: <http://feeds.delicious.com/v2/rss/username>

Sharing images, presentations and video *Flickr* for images, *Slideshare* for presentations, and *Vimeo* and/or *Youtube* for video. For high-profile items, use these services' capabilities to embed the results for viewing within the DCC Website.

6.5 Intellectual property policy

When soliciting contributions from our Associates Network we need similar procedures to those already established for the Curation Reference Manual, to ensure that contributors have IP rights to anything they contribute to a DCC resource, and that we can make the results available according to the Creative Commons 'non-commercial share-alike' licence.

6.6 Operating processes: using the platforms to further the objectives

The main body of the report has considered the various platforms and tools that our target communities use to converse about their activity, and our own. The tables below summarise how these might best be approached.

Note: these tables deal with DCC-initiated activity on the DCC Website and elsewhere, and with community activity elsewhere. They do not go into detail about what our target users for the Website should be able to do on that site, as the Web redevelopment project is considering that. The tables do however identify issues that may overlap. This section should be expanded to offer guidance on how to deploy the recommended tools/platforms.

DCC website actions	Parallel action on social networks	Issues
News: publish news item	<ol style="list-style-type: none"> 1. Tag news item on Delicious. 2. Shorten the item and url for microblogging and include link back to DCC site. 3. Expand on the item for a blog post if possible and include link back to DCC site. 4. Re-post the item to all services on which we have a corporate presence/account. 5. Consider doing same for individual accounts. 	<ul style="list-style-type: none"> • Approach implies all our accounts on selected platforms carry the same content by default. We can also elaborate for specific audiences by posting to relevant platforms or groups. • Quality assurance. • Confidence to express opinion, need to differentiate from news/information. • How to relate Delicious tags to DCC tags?

News: post blog item	<ol style="list-style-type: none"> 1. Tag blog post on Delicious. 2. Cross-post to DCC 'outposts'. 3. Consider search on Blogsearch and Technorati – do other blogs or forums have relevant posts you could comment on, based on your post? 4. Listen and respond. 	<ul style="list-style-type: none"> • Useful to identify specific high-ranking blogs we should target, to benefit from reciprocal impact on our own. • Quality assurance.
Event: call for participation	<ol style="list-style-type: none"> 1. Post to LinkedIn and/or Facebook. 2. Invite registered attendees to sign up to Associates Network. 	
Resource: publish resource	<ol style="list-style-type: none"> 1. Post news item on Website and cross-post to 'outposts'. 	
About: invite contact	<ol style="list-style-type: none"> 1. Invite Associates Network members to join us on Facebook, LinkedIn. 2. Invite connections/followers to join Associates Network. 	<ul style="list-style-type: none"> • Needs further consideration of what we can ask of Associates Network and 'champions'.

6.7 Implementation plan

- Address the issues highlighted in the report and begin implementing the recommendations.
- Consider roles and resources: who does what and where do they get the time?
- Locate or construct for each recommended tool relevant FAQs or 'how to' tutorials.
- Identify further deliverables.

A Survey e-mail text

How do social media platforms help you with research data management issues? At the Digital Curation Centre we are reviewing how we interact with social media beyond our own site at www.dcc.ac.uk. For any of the categories below that you have used, please tell us:

- Which platforms do you find valuable for research data issues?
- In a few words, what do you get out of using this platform?

List of categories

- Blogging/microblogging
- Online discussion/networking
- Sharing (links to) interesting Web resources
- Sharing (references to) interesting research papers, etc.

It would also help us if you could tell us your interest in research data. Please put an X against any that apply to you.

- Researcher
- Research Administrator
- Research Funder
- Data Manager/Curator
- Data Centre/Archive/Repository
- Library Service

Please send responses to a.ball@uoln.ac.uk rather than the list. We will circulate a summary of the responses ASAP.

B SurveyMonkey survey text

Introduction

At the Digital Curation Centre we are reviewing how we interact with social media beyond our own site at www.dcc.ac.uk. To help us with this, we would like to know how social media platforms help you with research data management issues.

We're particularly interested in the platforms you use for:

- Blogging/microblogging
- Online discussion/networking
- Sharing (links to) interesting Web resources
- Sharing (references to) interesting research papers, etc.

Blogging and microblogging

1. Which blogging/microblogging platforms do you find valuable for research data issues?
2. In a few words, what do you get out of using these platforms?

Online discussion/networking

1. Which online discussion/networking platforms do you find valuable for research data issues?
2. In a few words, what do you get out of using these platforms?

Sharing Web links

1. Which platforms for sharing links to interesting Web resources do you find valuable for research data issues?
2. In a few words, what do you get out of using these platforms?

Sharing research publications

1. Which platforms for sharing links to interesting papers, reports, etc. do you find valuable for research data issues?
2. In a few words, what do you get out of using these platforms?

Other social media platforms

1. What other social media platforms, if any, do you use to help you with research data management issues?
2. In a few words, what do you get out of using these platforms?

About you

It would also help us if you could tell us your interest in research data.

1. What roles do you play with respect to research data?
 - Researcher
 - Research Administrator
 - Research Funder
 - Data Manager/Curator
 - Data Centre/Archive/Repository
 - Library Service
 - IT Services
 - Other (please specify)

Bibliography

- Boulton, C. (19th May 2010). Google Wave now part of Google Apps, open to all. *EWeek: Enterprise Applications News*. Retrieved 21st June 2010, from <http://www.eweek.com/c/a/Enterprise-Applications/Google-Wave-Now-Part-of-Google-Apps-Open-to-All-111055/>
- Boulton, C. (8th June 2010). Google Wave This borrows from Twitter, Google Buzz. *Google Watch*. Retrieved 21st June 2010, from http://googlewatch.eweek.com/content/google_wave/_for_the_1_million.html
- Hutař, J., Melichar, M., Heister, C., Cubr, L. & Guttenbrunner, M. (31st March 2009). *Report on social networking presence for DPE, WePreserve and Digital Preservation/Curation (Deliverable DPE-D7.3)*. DigitalPreservationEurope. Retrieved 14th April 2010, from <http://www.digitalpreservationeurope.eu/publications/reports/D7.3ReportOnSocialNetworkingPresenceforDPE.pdf>
- Jadu. (2010, January). *An investigation into the challenges, application and benefits of social media in higher education institutions*. Retrieved 5th May 2010, from http://www.jadu.co.uk/download/18/research_into_the_challenges_usage_and_benefits_of_social_media_in_higher_education_institutions
- Joly, K. (2009, October). Social media measurement 101. *University Business*, 12(9). Retrieved 18th June 2010, from <http://www.universitybusiness.com/viewarticlepf.aspx?articleid=1416>
- Karthik, M. (30th March 2010). 10 social media monitoring/analytics tools for measuring social media. *Daily Bloggr*. Retrieved 21st June 2010, from <http://www.dailybloggr.com/2010/03/social-media-monitoring-analytics-tools/>
- LinkedIn. (18th June 2010). Retrieved from Wikipedia: <http://en.wikipedia.org/w/index.php?title=LinkedIn&oldid=368849479>
- McCarthy, C. (23rd June 2008). Facebook overtakes MySpace globally. *ZDNet News*. Retrieved 18th June 2010, from <http://www.zdnet.com/news/facebook-overtakes-myspace-globally/207724>
- Paul, I. (10th February 2010). Google Buzz versus Google Wave. *Today@PCWorld*. Retrieved 21st June 2010, from http://www.pcworld.com/article/189019/google_buzz_versus_google_wave.html
- Shiels, M. (14th April 2010). Congress to archive every tweet ever posted. *BBC News*. Retrieved 16th June 2010, from <http://news.bbc.co.uk/1/hi/technology/8621297.stm>
- Technorati Authority FAQ. (n.d.). Retrieved 18th June 2010, from Technorati: <http://technorati.com/what-is-technorati-authority/>