The Project Fragmentation Problem on the Web

David Jeffery
School of Information and Library Studies
University College Dublin
Dublin 4, Ireland
dave@davejeffery.com

ABSTRACT

An ever-increasing amount of our work, research and personal content are created online. We are living in a world of constant accumulation and the digital content we have created and curated online is dispersed across different devices, different online services, and created using different technologies and platforms.

The combination of an ever-expanding stream of personal information and the trend that sees more and more of this personal information being dispersed across different sources makes the accessibility of this content an important research issue. This paper investigates folders and activity streams as effective metaphors for organising project content. Furthermore, this paper details strategies for aggregating online personal content and abstracting the content into projects or collections. Finally, this paper will describe how OJAX++, an 'in-progress' VRE (Virtual Research Environment), has tackled the challenges presented in this paper.

1. INTRODUCTION

An ever-increasing proportion of our work, research and personal content are created online. We are living in a world of constant accumulation and the digital content we have created and curated online is dispersed across different devices, different online services, and created using different technologies and platforms.

As the cost of digital storage continues to fall, it becomes increasingly possible for us to record, store and track the content that we create and encounter online. In 2009, digital information grew by 62% to 800,000 petabytes (1 million gigabytes) according to The Guardian newspaper [15]. Evidence suggests that our capacity to mentally keep track of this information, however, is not increasing at the same rate [8]. Clutter may obscure valuable information and simply add to the burden of organising and managing the content; side effects of clutter include distracted attention and less effective searching or browsing [8].

The fragmented, dispersed nature of our content is a relatively new paradigm: in earlier times, we may have stored all of our content in a physical folder, and later, in a virtual folder on a hard drive. We are now beginning to store a large proportion of our content online, distributed across many different web services and web applications. Increasing use of tools like Delicious (http://www.delicious.com), Gmail (http://www.gmail.com), Flickr (http://flickr.com), Youtube (http://www.youtube.com) and Facebook (http://www.facebook.com) etc. have pushed more and more of our content and information onto the cloud and as our content is fragmented across an increasing number of different online applications, it's becoming increasingly less likely that we have a canonical folder, virtual or otherwise, where we can access all of the content related to a particular theme or project. This has forced users working on single projects to retrieve information from different sources, while having no semantic or structural connection between each source [2, 3, 10]. Research from Marshall and Jones [13] find that "the attendant fragmentation of our personal information increases the chances of keeping something in the wrong place or form and forgetting we ever saw, heard, or read it in the first place."

This "information fragmentation", increases cognitive load particularly when coupled with the inconsistent interaction designs of many web-based applications [3]. When working on a project it is typical for the user to set up a specific work context for the project, this may involve collecting and organising resources so that they are 'at hand' when working on that project [10, 11]. Users are constantly switching between different work contexts and this process can be time-consuming and difficult. A system for organising work contexts should have little overhead and allow the user to focus on doing work and achieving project goals rather than managing and organising their project data [3, 10].

In his now-iconic 1945 article "As we may think" [4]; Vannevar Bush focuses on the task of improving accessibility to our ever-increasing store of information and knowledge. Bush argues for the "privilege of forgetting the manifold things he does not need to have immediately, with some assurance that he can find them again if proven important". Luckily, we find that much of the content and information we accumulate online is archived for us by web applications. Simply having access to the information, however, is not enough, it must be abstracted and displayed in a useful and coherent manner if we are to interpret and use it properly. Abstracting this information is made difficult because of the fragmentation of online content, as discussed earlier, and the lack of a standardised format in which the data can be accessed.

iHCI 2010, September 2–3, 2010, Dublin, Ireland.

The combination of an ever-expanding stream of personal information and the trend that sees more and more of this personal information being dispersed across different sources makes the accessibility of this content an important research issue. This paper investigates folders and activity streams as effective metaphors for organising project content. Furthermore, this paper details strategies for aggregating online personal content and abstracting the content into projects or collections. Finally, this paper will describe how OJAX++ an 'in-progress' VRE (Virtual Research Environment), has tackled the challenges presented in this paper.

2. OJAX++

OJAX++ (http://www.ucd.ie/ojax/), is an 'in progress' VRE (Virtual Research Environment) focusing on the usability of online collaborative research. The project is funded by the SFI (Science Foundation Ireland) and is a collaboration between the UCD School of Information & Library Studies (SILS) and the UCD School of Computer Science & Informatics (CSI). OJAX++ aims to manage disparate content and knowledge from a wide variety of third- party tools.

The organisation of research and associated content in an integrated fashion across third- party tools is an ongoing research issue in the field of VREs. OJAX++ looks to address this issue. Its stated goal is to act as a single cohesive entity that provides the management and overview functions that would normally be associated with a monolithic, 'everything in one place' VRE.

The key features of OJAX++ centre around its ability to act as (1) a hub and central point for the aggregation of data from third party applications and (2) a communication and collaborative system for researchers wishing to collaborate using OJAX++.

3. FOLDERS

Creating, storing, recording, receiving and accumulating our online content is the easy part, managing them and storing them in a way that makes it easy to access them later is difficult. As mentioned in the introduction, the movement of our data away from personal hard drives and onto the web or the 'cloud' raises issues with regards to the findability or accessibility of that content. When all of our content was stored on our hard drive, the strategy was relatively straightforward; we might create a folder that related to a project we were working on and then simply add content to that folder. This strategy is not perfect, as we will explore later, but it is simple and easy to understand, and regardless of which desktop application the content was created in, it could be stored in that single location — a folder on your hard drive. The web is not a centralised store of data, if we want to use online applications then the metaphor of a catch-all project folder is largely irrelevant because our content is fragmented based on which application it was created in.

Folders are useful as more than simply a canonical store of a project's data. In a paper examining folders as an information management tool Jones et al. found that a hierarchy of folders and sub-folders often provided a rough outline for that project — over a period of time, folders represent a person's evolving understanding of a project and its components [9]. After being offered the alternative of a fast desktop search utility, Jones' study [9] found that out of the 14 participants only one answered that 'yes' — they would be willing to part with folder organisation. The paper concludes that the visibility and understandability benefits that a folder gives the user are seen as the primary reason for the reluctance to give up folder organisation.

The study did however highlight a number of issues with folders as an organisational tool; they found that users were prefixing folder names with characters like "aa" and "zz" to compensate for the lack of support for ordering. Interestingly, even though the user has complete control over the storage of data on their hard drive it is rare for all content relating to a project to be stored in a single folder, for example email messages and web bookmarks are generally handled in specific folders created by the source application [2]. Furthermore, some files may be related to several projects at once. Finally files are not the exclusive type of information object used in a project [10]. It's fair to summarise, therefore, that even when the user has complete control of the storage medium, some desktop applications can make it difficult for a user to store all their content in a single project folder.

4. ACTIVITY STREAMS

Freeman [7] in 1997 described activity streams as "...a time-ordered stream of documents that functions as a diary of your electronic life". On the web, activity streams come in a number of different guises and may be known as newsfeeds, lifestreams, digital traces, or 'social aggregation'.

Activity streams are a simple organisational metaphor, designed as an alternative or a replacement for the folders metaphor (discussed in the section 2). In an activity stream metaphor, all documents that you create are stored in your activity stream, alongside documents that other people send you. More recent documents are presented at the top of the activity stream, while older documents are towards the bottom. Activity streams can optionally also allow scope for future events such as reminders, schedules and to-do lists.

On comparing activity streams with the traditional folder metaphor, Freeman highlights the following benefits:

- Activity streams transparently store information, allowing users to concentrate on the task at hand rather than the name, folder, disk, machine, or network of a particular data item.
- Activity streams store information at the time it is created and organises information in the context it is needed. This reduces the overhead of creating information, improves recall, and facilitates retrieval.
- Activity streams provide new opportunities for users to exploit relationships and global patterns that exist in document collections by providing architectural framework for creating executive summaries and overviews.

4.1 Substreams

The metaphor can be further expanded using the concept of substreams [6]. Substreams allow dynamic filtering based on a given criteria to be performed on top of an existing activity stream. A typical substream filter might be: *"Bookmarks by John in July and August 2009, that contain the tag 'research'"*, this query simultaneously places constraints on actor, time, metadata and activity type; this gives the user a more concise view of the activity stream that is based on their current context needs. Figure 1, below, shows how OJAX++ applies substream functionality by allowing users to filter by tags, users and type; and sort by newest, oldest and most recently active. Substreams become particularly useful as projects or collections grow because:

- They give the user the ability to filter or order content in a collection to a narrower context.
- Larger collections of content offer more opportunities to correlate similarities based on common attributes.

Figure 1. Substream functionality in OJAX++. This substream would show items containing the tags 'collaboratory' or 'vre' by the user 'Dave J.', which are of type 'Note', 'Bookmark' or 'Activity'; these are then sorted in chronological order (oldest first).

5. AN APPROACH TO SUPPORTING HIGHER-LEVEL USER ACTIVITIES ON THE WEB

Folders and Activity Streams are both effective organisational structures for managing projects and their associated content. On the web, however, their usefulness is limited because project activity is fragmented by application and *not* by theme or project. The web requires a different approach to supporting higher-level user activities such as projects on the web.

The following approach describes a method that (a) creates an information bucket (or project) containing related items from (b) different applications spread across the web, utilising a (c) transparent process for adding project data that (d) allows for support of additional web applications through a plug-in architecture. OJAX++ uses a two-step process to achieve support of higher-level user activities (projects) across multiple web applications:

1. Aggregate all of the project content across source applications.
2. Based on common attributes, abstract this content into projects.

5.1 Aggregation

A number of aggregators of personal online content already exist, the most popular is FriendFeed (http://www.friendfeed.com), which aggregates personal content from 58 sites and presents the aggregated content on an activity stream. To successfully aggregate content, it is important to normalise the data so that it is not in a form that is specific to the source application. For example, a bookmark should have common fields regardless of its source, e.g. Delicious or Diigo (http://www.diigo.com). Dedicated Services such as GNIP (http://www.gnip.com) are created to specifically solve this issue for developers wishing to create applications that aggregate personal online content.

Aggregation of content from third-party applications relies on the

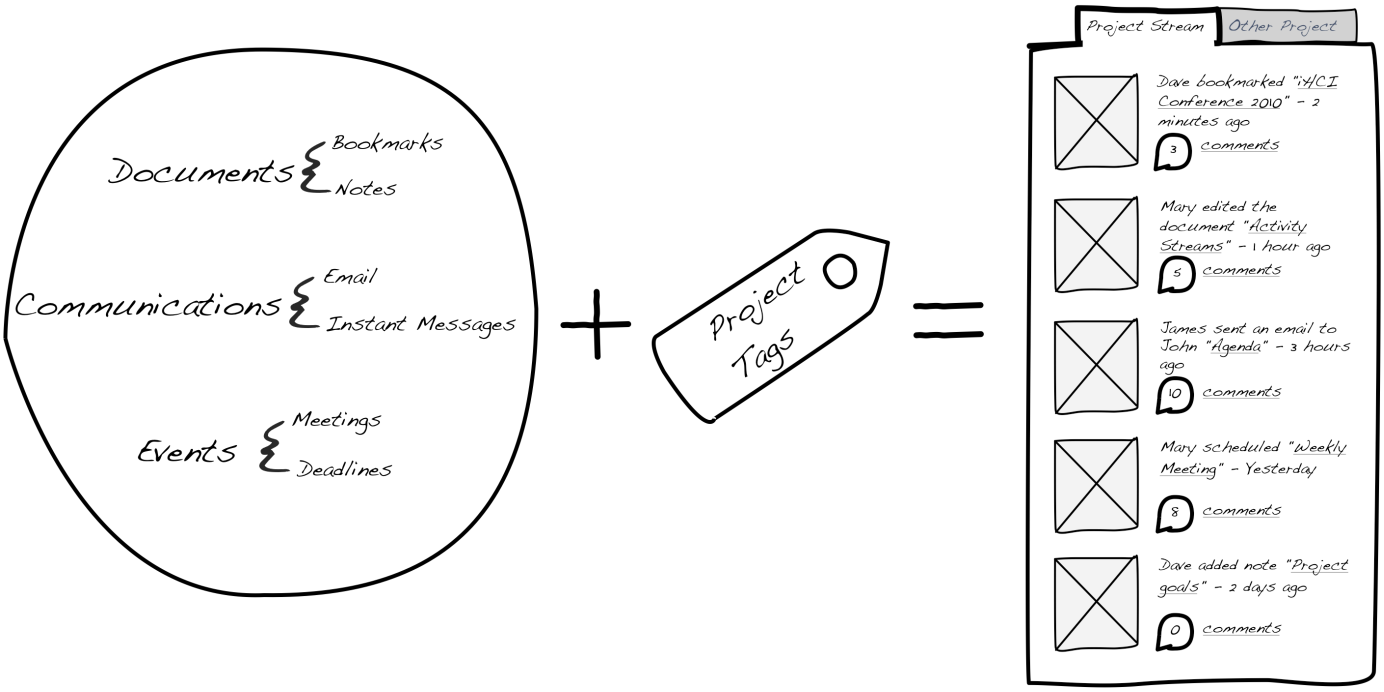

Figure 2. Tagging can be used as an abstraction technique to categorise content into projects as described in section 5.2.

availability of a public API (Application Programming Interface) to allow for ad-hoc integration with other applications. The proliferation of public APIs is a relatively recent phenomenon, and coupled with the continued adoption of standards-based approaches, it provides a scalable solution for aggregating content from a number of different providers. Another method for aggregating content is 'screen-scraping' but this method is more difficult for developers to implement and also unreliable because webpage layouts may change frequently and may brake screen-scraping algorithms [5].

OJAX++ provides a plug-in architecture for aggregating content from third party websites and mapping the content to a normalised model. A plug-in architecture allows additional sources to be added easily in the future, assuming that they have a public API.

5.2 Abstraction

Abstraction of content into projects could be done either manually or automatically. A manual solution may involve interaction from the user asking them to specify which project(s) they wish each item of content to be associated with. An automatic abstraction solution may involve techniques such as data mining or examining common meta-data, assuming that it has been normalised during the aggregation process.

The rise of 'Web 2.0' style applications has seen a surge in the popularity of 'tagging' as a strategy for categorising content [12]. Tagging, therefore, is an ideal candidate for abstracting content into projects (see figure 2). OJAX++ uses tagging as a common classification scheme to denote the target project of an item of content. In OJAX++, a user creates a project and chooses tags which become 'project tags' — when content from a third-party application is tagged using one of the 'project tags' it will automatically be added to the project stream in OJAX++.

6. DISCUSSION

In the process of researching and drafting this paper: I bookmarked and annotated interesting articles and papers using the online tools Delicious and Diigo. I archived my hand-written notes using an application called Evernote on my mobile phone. I saved relevant research papers as PDFs on Dropbox, a backup and synchronisation tool. I discussed the paper with my supervisor using Gmail and I drafted the paper using Google Docs. In total, personal content relating to this paper is stored across six different online services and there is no semantic or structural connection between each service that allows me to group this data into one project.

Research has shown that users prefer to relate their information to higher-level user activities (i.e. projects) [1, 10], however, the decentralised design of the web implicitly discourages this and encourages project fragmentation. Continued adoption of standards-based development practices provide a clear path for an intermediary application that aggregates activity across a range of services and abstracts higher-level user activities. OJAX++ is an application that takes this approach and utilises user-defined 'tags' as a method for defining and organising project activity.

7. ACKNOWLEDGMENTS

Thanks to the OJAX++ team; in particular I thank Judith Wusteman, the principal investigator of the OJAX++ project. This work is funded by the Science Foundation Ireland (SFI) and is a collaboration between the School of Information and Library Studies (SILS) and Computer Science and Informatics (CSI) departments in UCD.

8. REFERENCES

[1] Bergman, O., Beyth-Marom, R., Nachmias, R. 2006. The Project Fragmentation Problem in Personal Information Management. In CHI'06 Proceedings on Personal Information Management (Montréal, Québec, Canada, Apr. 22–27) 247-274.

[2] Boardman, R. Sasse, M.A., 2004. "Stuff goes into the computer and doesn't come out" A Cross-tool Study of Personal Information Management. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2004).

[3] Boardman, R., Spence, R., Sasse, M.A. 2003. Too Many Hierarchies? The daily struggle for control of the workspace. In HCI International 2003, Lawrence Erlbaum. 626-620.

[4] Bush, V. 1945. As we may think. The Atlantic Monthly 176, 1 (July 1945).

[5] Daconta M., Obrst L., Smith K. 2009. The Semantic Web: a guide to the future of XML, Web services, and knowledge management. Wiley Publishing.

[6] Fertig S., Freeman E.T., Gelerntner D. 1996. "Finding and Reminding" Reconsidered. SIGCHI 28 1.

[7] Freeman, E. T. 1997. The lifestreams software architecture. PhD thesis. Yale University, Connecticut, United States.

[8] Jones, W. 2004. Finders, keepers? The present and future perfect in support of personal information management. First Monday 9, 3 (Mar. 1, 2004), http://www.firstmonday.org/issues/issue9_3/jones/.

[9] Jones, W., Phuwanartnurak, A., Gill, R., and Bruce, H. 2005. Don't take my folders away! Organizing personal information to get things done. In CHI'05 Extended Abstracts on Human Factors in Computing Systems (Portland, OR, USA. Apr. 2–7). ACM Press, New York, 1505–1508.

[10] Kaptelinin V. 2003. UMEA: translating interaction histories into project contexts. Proceedings of the SIGCHI conference on Human Computer Interaction 5 1.

[11] Malone, T.W. 1983. How do people organise their desks? Implications for the design of office information systems. ACM Transactions on Office Information 1.

[12] Marshall, C. C. 2009. No Bull No Spin: A comparison of tags with other forms of user metadata. In JCDL '09 proceedings (Austin, TX, USA. Jun. 15–19)

[13] Marshall C. C., and Jones W. 2006. Keeping Encountered Information. Communications of the ACM 49,1 (Jan. 2006).

[14] Nardi, B. and Barreau, D.K. "Finding and reminding" revisited: appropriate metaphors for file organization at the desktop. SIGCHI 29 1.

[15] Wray, R. 2010. Goodbye petabytes, hello zettabytes. The Guardian, (May 3 2010), http://www.guardian.co.uk/technology/2010/may/03/humanity-digital-output-zettabyte.

[16] Wusteman, J. 2008. Virtual Research Environments: What is the Librarian's Role?. Journal of Librarianship and Information Science 40 2 67-70.