Patterns of Usage in an Enterprise File-Sharing Service: Publicizing, Discovering, and Telling the News

Michael Muller, David R Millen, & Jonathan Feinberg
IBM Research / Collaborative User Experience & IBM Center for Social Software
One Rogers Street, Cambridge, MA, USA 02142
{michael_muller, david_r_millen, jdf} @ us.ibm.com

ABSTRACT
How do people use an enterprise file-sharing service? We describe patterns of usage in a social file-sharing service that was deployed in a large multinational enterprise. Factor analyses revealed four factors: **Upload & Publicize** (regarding one's own files); **Annotate & Watch** (add information to files and maintain awareness); **Discover & Tell** (find files uploaded by other users, and communicate to additional users about those files); and **Refind** (re-use one's own files). We explore the attributes of users who score highly on each of these factors, and we propose implications for design to encourage innovation in usage.

Author Keywords
Social-software, file-sharing, enterprise.

ACM Classification Keywords
H5.3. Group & Organizational Interfaces: CSCW.

General Terms
Human Factors.

INTRODUCTION

In the process of adopting a new technology, users often adapt the technology or their working practices [2,4,7,11]. This short note describes the emergence of four usage patterns in Cattail, an experimental file-sharing service in IBM. *Anticipated* patterns included core components of sharing and re-use. *Unanticipated* patterns included core components of publicizing one's own files and sharing one's discoveries.

Cattail was originally designed to provide an alternative to sending large attachments via email (e.g., [12]). It is in some ways similar to the centralized sharing services studied by Rader [10] and Voida et al. [12], but has been designed with features to support large enterprise use. It is a centralized service, requiring full user authentication, and thus avoids some of the issues that have arisen in peer-to-peer and non-authenticated sharing services, as reported by [7]. Based on user-authentication, the design evolved to include many social features [7], and these features led to new working practices. This note reports emergent working practices in a large-scale quantitative study.

The remainder of this note is organized as follows. We begin with a high-level description of Cattail. We describe the data used in our analyses. We present then a factor analysis of user activities, resulting in four factors. Next, we seek to understand those factors by looking at the attributes of users who scored highly on each of those factors. We close with a discussion of social-software systems as sites for innovative use, and implications for design.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.
Copyright 2010 ACM 978-1-60558-929-9/10/04....$10.00.

CATTAIL FEATURES AND CAPABILITIES

A description of the Cattail user experience appeared in [7]. In this note, we focus on the subset of features that are relevant to the phenomena in our factor analyses.

The user may upload files, and may download any file to which s/he has access – i.e., all files classified as "Public", plus those files which s/he had Uploaded, plus limited-access ("Private" and "Confidential") files that had been Shared to her/him. Users are required to make an explicit choice of the access category during the Upload operation – i.e., there is no default, and the upload cannot continue until the user makes this selection.

The operation of Sharing involves granting access to one or more other users, and sending a notification to those people. The user may Share any file to which s/he has access.

The user may also create a named Collection, and may add files to that Collection. Collections function as labels rather than containers: Any file may appear in zero, one, or many Collections. Collections may be accessed by other users [7].

The user may create an Annotation (a comment) on a file. Annotations may be read by anyone who has access to the file on which the Annotation was written.

METHOD

For this analysis, we used a dataset of all actions in Cattail from its first date of operation (17 May 2007) through 10 December 2008.

Upload: number of files Uploaded by that user (15933 users, or 18.1%)

Download-Own: number of files Downloaded by the user *which had previously been uploaded by that user* (5149 users/5.8%)

Download-Other: number of files Downloaded by the user *which had previously been uploaded by other users* (85188 users/96.5%)

Share-Own: number of files Shared by the user *which had previously been uploaded by that user* (11400 users/12.9%)

Share-Other: number of files Shared by the user *which had previously been uploaded by other users* (2838 users/3.2%)

Collect-Own: number of files placed in Collections by the user *which had previously been uploaded by that user* (4711 users/5.3%)

Collect-Other: number of files placed in Collections by the user *which had previously been uploaded by other users* (1720 users/1.9%)

Annotate-Own: number of files on which the user wrote comments, *which had previously been uploaded by that user* (1309 users/1.5%)

Annotate-Other: number of files on which the user wrote comments, *which had previously been uploaded by other users* (2977 users/3.4%)

Watch-Other: number of files for which the user requested email notifications, *which had previously been uploaded by other users* * (6170/7.0%)

* The category of Watch-Own was omitted from the analysis, because only one user employed this option.

Table 1. The ten user activities in the factor analysis, and the number of users exhibiting each of those activities.

The service was offered to all of the employees of IBM, including employees in areas of software, hardware, consulting, and research as well as internal operations in support of those external-facing activities. The service was not advertised to employees, but many employees try out new experimental services as part of IBM's innovation process.

As of the close of the study period, 88270 people had used Cattail at least once, and 15934 of them had uploaded at least once file (this ratio is somewhat *less* lopsided than similar results from the public internet [5,8]). Users came from many parts of the company and from many job roles, including over 80 countries. Over the 18 months of our study, 120288 files were uploaded, and a total of 728509 download operations occurred on those files.

FACTOR ANALYSIS RESULTS

Factor analysis is a useful statistical method for "data reduction" when there are many variables of potential interest. In our case, we found ten core activities that might be performed by users of Cattail (Table 1). Of course, these activities were inter-correlated. We used factor analysis to understand the pattern of inter-correlations, and to identify a smaller number of inter-related activities that tended to co-occur from one user to another. Thus, we used factor analysis to "reduce the data" from the original ten activities to a smaller number of "factors" of highly-related activities.

Specifically, we performed a series of principal components factor analyses. Each user constituted one observation in the analysis, characterized in terms of the ten activities described in Table 1. Table 2 presents a four-factor solution with Varimax rotation, accounting for 69% of the variance. In factor analysis, each factor (combination of variables) is associated with an eigenvalue. Typically, factors that have eigenvalues greater than 1.0 are of interest, while factors with lower eigenvalues are usually ignored. Interpretation is then based on the weight (degree of association) of each activity with each factor (the values in the cells of Table 2).

Upload & Publicize: The first factor had high loadings for the activities of Upload of files, Share-Own files, and Collect-Own files. These three activities form a straightforward "package" of actions through which a user can upload factors for use by other people – i.e., aggregating relevant files into named Collections, and providing access (Sharing) those files with other users.

	Factor			
Eigenvalue	2.995	1.192	1.061	1.001
%Variance	29.945	11.919	10.610	10.011
Measure	**Upload & Publicize**	**Annotate & Watch**	**Discover & Tell**	**Refind**
Upload	• **.935** •	.100	.125	.063
Download-Own	.064	.021	.014	• **.992** •
Download-Others'	.033	.066	• **.858** •	.042
Share-Own	• **.734** •	.202	.198	.073
Share-Others'	.241	.175	• **.563** •	-.015
Collect-Own	• **.855** •	.062	.035	-.028
Collect-Others'	.085	• **.523** •	• **.407** •	-.048
Annotate-Own	.276	• **.756** •	-.021	.071
Annotate-Others'	.134	• **.874** •	.097	.046
Watch-Others'	.025	• **.761** •	.184	-.040

Table 2. Factor loadings for 88270 users, principal component analysis with Varimax rotation (69.115% of variance). We considered any cell (association between an activity and a factor) with a value > .400.

This factor is consistent with results of [1,7,11], which reported social-software as means to reach an "audience" of other users.

Annotate & Watch: The second factor had high loadings for the activities of Annotate-Own files, Annotate-Others' files, and Watch-Others' files, which suggests that people were adding information to files, and were remaining aware of activities related to those files. Previous research has shown the importance of awareness of events regarding persons and objects in social-software systems [6,9,10,12].

Discover & Tell: The third factor had high loadings for the activities of Download-Others' files, Share-Others' files, and Collect-Others' files. This factor is a complement to the first factor. Users in the **Upload & Publicize** factor appeared to be making their *own* files useful to other people, whereas users in the **Discover & Tell** factor appeared to be doing similar work with files that had been uploaded by *other people*. As noted earlier, this kind of work is consistent with reports of users' intentions of sharing resources with "audiences" of other people [1,7,11].

Refind: The fourth factor had a high loading for Download-Own files. This factor appeared to reflect people using Cattail as an extension of their own harddrives.[1]

DISCUSSION

The four factors described in this paper may be considered as representative of the work practices of many users in an enterprise social-software file-sharing environment. These work practices offer a lens onto ways in which users adopt and adapt technologies into their work (see also [2,4,7,11]). Our results show that users have made Cattail both more social and less social than was originally anticipated.

More Social

We interpret the social results through the lens of a qualitative study of "curators" in Cattail (i.e., users who deliberately Collect and Share files for use by other users) [7h] and three other studies of file-sharing [9,10,12]. Our factor analysis was consistent with the "curators" finding: People Shared files broadly, and constructed Collections of files for use by other people (**Upload & Publicize** and **Discover & Tell**). Some users sought out files that had been contributed by others, Collected those files, and reShared those files to yet other users (**Discover & Tell**). This widespread sharing led to interesting summary statistics that 44395 (37%) of the all of the files in Cattail were stored as Public (readable by anyone in the company); 60476 (50%) of the files had been Collected; and 63010 (52%) of the files had been Shared.

However, unlike [9,12], Cattail users were likely to share *work*-oriented rather than *personal* files. Unlike [10], the Cattail users worked in a very large company, where sharing could occur with people who were not personally known to the user. Indeed, one emergent outcome of enterprise social software is the discovery of previously unknown collaborators through the sharing system [6].

This quantitative approach can expand and refine concepts from more qualitative analyses. Based on interviews, [7] reported a usage pattern of "information curators" who assemble groups of files for use by other users. Our analyses suggest that the curating in fact occurs in two distinct patterns of use, i.e., calling attention to a selected subset of one's own files (**Upload & Publicize**) vs. communicating about a selected subset of others' files (**Discover & Tell**).

In contrast to the problems critiqued in [10], extensive sharing took place in Cattail *without* impacting the content of the files or the ability of an Uploader to Refind her/his files. Users promoted the files to other users' awareness (Sharing), and enhanced the available information about those files (Collecting, Annotating). One solution to the sharing problems in [10] is to leave the files in place, but provide social engagement mechanisms *around* those files.

Less Social

This note also shows the importance of Cattail for an unanticipated pattern of adaptation, which was the use of Cattail as an extension of the user's hard drive (i.e., the **Refind** factor), despite the fact that IBM provides many opportunities for back-up and storage *outside* of the social software domain. In total, 16994 files (14%) were used in this way by 5149 people (32% of Uploaders). The fact that many of these Refound files had "Public" access permissions is puzzling. (We note that, as stated above, users were required to specify one of the three access permission categories during an upload operation; there was no default permission. Thus, each permission status was an explicit user choice.) Perhaps the Public Refound files constitute a *failure case*, in which the Uploader hoped for an audience but failed to interest anyone in becoming that audience.

As argued in [2,4], users often construct new and sometimes surprising meanings and work practices to accommodate new technologies into existing activities. These user-initiated insights can provide the basis for new features and enhanced services. Social software offers a fertile setting for such user innovations (e.g., [2,4,7,11]). The design of social software to encourage user innovations during adoption, is a promising area for new research.

[1] Reviewers have asked whether **Refind** is merely an uninformative, "all-remaining variance" outcome. We also explored three-factor and five-factor solutions. These solutions included a similar **Refind** factor. We note, as well, that the eigenvalue for **Refind** is just over 1.0, and that it accounts for roughly the same amount of variance as two the other four factors in Table 2.

Factor	Anno. & Watch	Disc. & Tell	Refind	Totals
Upload & Publ.	4301 (4.9%)	13269 (15.0%)	5143 (5.8%)	15935 (18.1%)
Anno. & Watch	-	9368 (10.6%)	2054 (2.3%)	9780 (11.1%)
Disc. & Tell		-	4640 (5.3%)	85303 (96.7%)
Refind			-	5149 (5.8%)

Table 3. Number of people associated with each factor, and with each factor *x* factor combination.

Implications for Design

An Omnibus User Interface, or Four Specialized Interfaces?

Our factor analysis showed four distinct patterns of usage. It is reasonable to ask, *Should each usage pattern have its own specialized user interface features?* One approach to this question is to examine the overlap among users who engaged in each of the four patterns. Factor analysis does not provide an explicit classification procedure. Therefore, we examined this question in several post-analysis steps.

First, for each user, we coded each of the ten user activity variables in Table 1 as a binary (either the user had performed the activity at least once, or s/he had not done so). Second, for each factor for each user, we summed the binary activity scores for the subset of activities that were associated with each factor (the highlighted cells in Table 2). For example, the summed score for the **Upload & Publicize** factor was ***Binary(Upload)*** + ***Binary(Share-Own)*** + ***Binary(Collect-Own)***. Finally, we recoded each summed score as a binary (the "Totals" column of Table 3). The **Discover & Tell** factor described the largest number of users (96.7%), although the **Upload & Publicize** factor also accounted for nearly a fifth of all of the users. **Refind** accounted for a comparatively smaller number of users.

To answer the question of *different interfaces?*, we examined the overlap among the binary scores for each two-way combination of the four factors (Table 3). The smallest overlap was 2054 (2.3% of the users). The largest overlap was 13269 people (15.0%). The mean overlap was 6265.5 users (7.3%). These results show that thousands of users engage in more than one role (as defined by the four factors). Therefore, we propose that a single, unified user interface should be provided, with features to support transitions among the different activities and roles (factors).

Features to Support Each Role

Users who have adopted the **Upload & Publicize** pattern may want better tools for publicizing, especially the ability to promote not only individual files to other users, but also to promote Collections of files (promotion of Collections is not a Cattail feature). These users may also be treating Cattail as an opportunity for reputation management [3]. These users may want to know not only whether their files have been accessed by others (a current feature in Cattail), but also how well those files were received. A rating system for files and for Collections may be useful to these users.

We hypothesize that users who have adopted the **Discover & Tell** pattern may have similar needs, especially regarding the ability to promote both individual files and Collections of files. Because both of these groups of users show patterns of sharing with others, they may also benefit from knowledge of the search terms used by other users (see also [11]). This capability is not currently available in Cattail.

We suspect that the users who have adopted the **Annotate & Watch** pattern may benefit from more structured discussion threads for each file, and they may be particularly interested in knowing whether their annotations have been read by other users (not supported in Cattail).

Finally, the users who have adopted the **Refind** pattern may want enhanced abilities to compare and synchronize files between their personal machines and Cattail (not currently supported). They may also benefit from the ability to filter their searches to examine only their own files.

REFERENCES

1. Ames, M. & Naaman, M. Why we tag: motivations for annotation in mobile and online media. *Proc.CHI 2007*.
2. Dourish, P., Process descriptions as organisational accounting devices: The dual use of workflow technologies. *Proc. GROUP 2001*.
3. Goffman, E., *The presentation of self in everyday life*, Doubleday, New York, 1959.
4. Kujala, S., & Kauppinen, M., Identifying and selecting users for user-centered design. *Proc NordCHI 2004*.
5. Li, C., & Bernoff, J., *Groundswell: Winning in a world transformed by social technologies*, Harvard Business School Press, 2008.
6. Millen, D.R., Yang, M., Whittaker, S., & Feinberg, J., Social bookmarking and exploratory search. *Proc ECSCW 2007*.
7. Muller, M., Millen, D.R., & Feinberg, J., Information curators in an enterprise file-sharing service. *Proc. ECSCW 2009*.
8. Nielsen, J., 'Participation inequality: Encouraging more users to contribute,' http://www.useit.com/alertbox/participation_inequality.html, 2006.
9. Olson, J.S., Grudin, J., & Horvitz, E. Toward understanding preferences for sharing and privacy. *CHI 2005 Extended Abstracts*.
10. Rader, E. Yours, mine and (not) ours: Social influences on group information repositories. *Proc CHI 2009*.
11. Thom-Santelli, J., Muller, M., & Millen, D.R., Social tagging roles: Publishers, evangelists, leaders. *Proc CHI 2008*.
12. Voida, S., Edwards, W.K., Newman, M.W., Grinter, R.E., & Ducheneaut, N. Share and share alike: Exploring the user interface affordances of file sharing. *Proc CHI 2006*.