

# IBM Research Report

## Patterns of Tag Usage across Four Diverse Enterprise Tagging Services

**Michael J. Muller**  
IBM Research Division  
One Rogers Street  
Cambridge, MA 02142 USA



**Research Division**  
Almaden - Austin - Beijing - Cambridge - Haifa - India - T. J. Watson - Tokyo - Zurich

# Patterns of Tag Usage across Four Diverse Enterprise Tagging Services

Michael J. Muller

IBM Research / Collaborative User Experience  
One Rogers Street  
Cambridge, MA, USA 02142  
michael\_muller@us.ibm.com

**Abstract:** We compare four tagging-based enterprise services, that respectively stored bookmarks to web-pages and documents, to people, to blog entries, and to hierarchically-structured activity records. Analysis of user data and tag data showed diffuse patterns of commonalities and differences across the services. These results will help us to understand emergent work practices, and the value of tagging services to enterprises.

In our research on social computing within the enterprise, we are concerned to understand how people interpret and use tags in different contexts. Tagging services within an enterprise can take advantage of full authentication of users, facilitating the identification of experts [3,5], people with shared interests [6], features of a shared vocabulary [1], and colleague-based strategies of information search [6].

Full authentication also makes it possible to compare users' tagging practices in multiple tagging-based services. This paper goes beyond the previous research, by examining tagging behaviors across four distinct tagging-based services, with partially-overlapping user populations, within a single enterprise. This paper takes a comparative view that was not possible in previously published single-service analyses.

Damianos et al. [1] described a pattern of vocabulary stabilization with increasing tag usage (see also [6]). We hope to learn if people expect each tag to have a stable, consistent meaning across services, and if their tag usage is consistent across the multiple types of objects stored in the multiple services. These studies will help us to understand the semantics of social software, which will in turn help us to understand the value of social software services to enterprises.

We collected tag data from the four services summarized in Table 1. **Dogear** is a social-tagging service for resources

such as public URLs, company-internal URLs, and other company internal documents (e.g., Wiki pages, Domino documents, etc.) [6]. Dogear bookmarks are readable by anyone in the company. **Bluepages+1** is an enhanced version of the company's online employee directory, in which one person may apply a tag directly to another person's directory page [2, 3]. Bluepages+1 data (including tags) are readable by anyone in the company. **Blog Central** is an internal blogging service, open to any employee. All entries in Blog Central are readable by anyone in the company. Finally, **Activities** [7] is a web-based version of ActivityExplorer [8], an activity-centric collaboration service in which teams may create a collection of diverse objects in a tree-like structure consisting of a root "activity" and its daughter components. Unlike the other three services, Activities restricts access to each activity to a specified access control list, which may be as small as one person, or as large as several hundred people.

For the purposes of this paper, data consisted primarily of user-tag tuples. Note that previous studies of tagging behaviors *within a single service* have included not only user and tag (as in our study), but also the tagged object [1, 2, 3, 4, 5, 6]. By contrast, this paper is concerned with tagging *across diverse services*, where each service is concerned with different types of tagged objects (e.g., documents vs. people vs. blogs vs. activity components). The same objects cannot appear in different services, because each service stores a different class of objects. Therefore, for this initial report, we omit references to the tagged objects, and focus only on the users and the tags.

Analyses showed complex patterns of overlaps. Many of the 4987 unique users participated in more than one of the services. On a pairwise basis, the number of people writ-

System	What is tagged	Number of Taggers	Number of Unique Tags	Potential Number of Readers
<b>Dogear</b>	Public URLs, internal URLs, and other internal documents	1710	21072	As many as 300,000 IBM employees
<b>Bluepages+1</b>	Persons	713	2992	As many as 300,000 IBM employees
<b>Blog Central</b>	Blog entries (tagged by author only)	2092	3322	As many as 300,000 IBM employees
<b>Activities</b>	Activity root object + daughter components of the activity	742	3269	From 0 to several hundred

Table 1. The four internal tagging services in this study. Data were collected from Dogear, Bluepages+1, and Activities on 21 July 2006. Data were collected from Blog Central on 9 August 2006. Usage has increased since those dates.

ing tags into each pair of services ranged from 193 users (Bluepages+1 and Blog Central) to 613 users (Dogear and Activities). Smaller numbers of users participated in more complex intersections of the services, with only 79 of the users participating in all four services.

When we examined tags in the four services, we found similar complex patterns. Of the 28460 unique tags, many appeared in more than one service. On a pairwise basis, the number of tags appearing in each pair of services ranged from 522 tags (Bluepages+1 & Activities) to 2953 tags (Dogear & Blog Central). Smaller numbers of tags appeared in more complex intersections of the services, with only 396 tags appearing in all four services.

Tags occurred with different relative frequencies in different services. Using the Fisher r-to-z transformations, we calculated the mean correlation of relative tag frequencies across services as  $r=.286$ . While significant ( $p<.001$ ), this correlation explains only 8% of the variance. People appear to use tags in somewhat different ways in the different services.

In addition to comparing lists of users and tags on a *per-service* basis (above), we also conducted analyses at the level of each person and each tag. When we focused on personal re-use of tags across services, we found very low re-use rates (mean of 1.06 tags re-used per person, or mean of 2.65% of per-user opportunities to have tags in common). When we re-focused our analysis on the tags, and asked whether each tag was associated with the same users across services (a relationship that we called “tag membership”), we again found very low rates of commonality (mean of 0.35 people associated with the same tag in different services, or mean of 4.89% of per-tag opportunities to have users in common).<sup>1</sup>

It could be argued that these results are not very surprising: If people were describing different types of objects, then shouldn't they use different vocabularies to do so? This explanation is not supported by the data. If people used different vocabularies, then we should have found very few tags in common between services. However, we found a

---

<sup>1</sup> We considered that these results might have been due to a sampling mistake, i.e., the mistake of including people who might have created only a very small number of tags in a service. To test this sampling-error hypothesis, we sorted each set of data in terms of the minimum number of tags created by the user in *each* of the services under analysis. We then experimented with temporarily removing the users with the lowest numbers of tags in one service or the other. We systematically tried all possible cut-off usage rates. None of the cut-off rates substantially increased the re-use of tags by persons. None of the cut-off rates substantially increased the number of people associated with each tag. The phenomena of low tag-re-use and low tag-membership appear to be robust across users with *any* number of tags.

minimum of 522 tags in common across pairs of services. Thus, the re-use of the tags, on a *service-by-service* basis, provided ample opportunity for the re-use of tags on a *person-by-person* basis. Nonetheless, we found very little personal re-use of tags. We also note that, at a conceptual basis, all four services were designed to support people's work, and therefore it appears reasonable to expect people to use similar vocabularies to describe the related aspects of their work that involved different types of objects in different tag-based services.

Other research within two of these services has shown that people devote considerable effort into writing tags [6], and that in some cases they use the tags as part of managing their reputations within the enterprise [3]. We are left with the apparent paradox that people work hard to write tags, but seem not to do this consistently for related aspects of their work that are stored in different services.

Future analyses will attempt to explain this apparent paradox through more a detailed comparison of the tags that do and that do not involve tag re-use and tag membership across services. Future analyses will also examine co-occurrence and network analyses of tags and taggers, as well as the semantics of the tags. We anticipate combining these analyses with other examinations of tagging patterns across diverse services and resources, to understand the emerging tag-based work practices, and to clarify the value of social software data and services within the enterprise.

## References

- [1] Damianos, L., Griffith, J., & Cuomo, D., “Onomi: Social Bookmarking on a Corporate Intranet,” *Proc WWW 2006*.
- [2] Farrell, S., & Lau, T., “Fringe Contacts: People-Tagging for the Enterprise.” *Proc WWW 2006*.
- [3] Farrell, S., Lau, T., Muller, M.J., & Ehrlich, K., “Augmenting employee profiles with people-tagging,” Submitted to CHI 2007.
- [4] Golder, S.A., & Huberman, B.A., “Structure of Collaborative Tagging Systems,” *J. Info. Sci.* 32, 2 (April, 2006).
- [5] John, A., & Seligmann, D., “Collaborative Tagging and Expertise in the Enterprise,” in *Proc WWW 2006*.
- [6] Millen, D.R., Feinberg, J., & Kerr, B., “Dogear: Social Bookmarking in the Enterprise,” *Proc CHI 2006*.
- [7] Moore, M., Estrada, M., Finley, T., Muller, M.J., & Geyer, W., “Next generation activity-centric computing. Demo at CSCW 2006.
- [8] Muller, M., Geyer, W., Brownholtz, B., Wilcox, E., & Millen, D.R., “One Hundred Days in an Activity-Centric Collaboration Environment based on Shared Objects,” *Proc. CHI 2004*.