Public vs. Publicized: Content Use Trends and Privacy Expectations

Jessica Staddon
Google
staddon@google.com

Andrew Swerdlow
Google
aswerdlow@google.com

Abstract

From a semantic standpoint, there is a clear differentiation between the meanings of public and publicized content. The former includes any content that is accessible by anyone, while the latter emphasizes visibility – publicized content is actively made available. As a user's online experience becomes more personalized and data is increasingly pushed rather than pulled, the line between public and publicized content is inevitably blurred. In this position paper, we present quantitative evidence that despite this trend, in some settings users do not anticipate the use of public content beyond the narrow context in which is was disclosed; they do not anticipate that the content may be publicized. While providing a "publicized" option for data is an important counterpart to the ability to limit access to data (e.g. through access control lists), such an option must be accompanied by both greater user awareness of the ramifications of such an option and by transparency into data usage.

1   Introduction

There is a consistent trend toward personalization of online content. As examples, users routinely receive search results personalized based on their online behaviors and content is often prioritized via the recommendations of social connections (e.g. Facebook personalized sites [8]) or users with similar preferences (e.g. Google News [4]). When personalization works well, and increasing adoption indicates it often does, users receive more relevant, interesting content than they would through nonpersonalized services.

Personalization clearly comes with privacy concerns because it requires analysis of how a user or browser interacts with services over time. In particular, to make recommendations based on behaviors and preferences of users, the service has to have this information, and to enhance these recommendations with social information,

the service must also have social connections. Hence, personalization further fuels the content needs of service providers. At a minimum it motivates the use of *public* and *semi-public* content (i.e. content that is public within the walls of a service or a large social network) to support personalization, consequently blurring the line between public and *publicized*.[1] For example, while comments on a blog may be public, when the blog is enabled with Facebook comments [5], a logged-in user's comments are pushed to their Facebook stream automatically, hence they are *publicized* to a degree. Indeed, publicizing of public and semi-public content is common in online social networks; profile changes by default trigger email notifications to contacts on LinkedIn, and by default Facebook friends receive notifications about many of their friends' game-related activities in Facebook [7].

While users should obviously exercise caution when allowing content to be publicized, we argue that having the ability to "publicize" content is as important as the ability to control access to content. User privacy preferences are often complex (see, for example, [1, 3, 26]) and settings that allow users to express a range of content delivery options (including publicize, "lock down", and options in between) are more capable of accurately reflecting preferences. In addition, given that an online provider may be quite aggressive in their use of use public content (e.g. Spokeo [28], which gathers the hobbies, property value, age and marital status of an individual from public records) and public acts or content may even be publicized by individual users [10], from a privacy point, it is far safer for the user to equate the two.

Our main contribution in this short paper is quantitative evidence that in some settings user expectations are not compatible with content use trends. We present studies with 380 users across the U.S., Germany and Italy,

---

[1]For simplicity, we predominantly use "public" to refer to content that is public to the world, to a service or a to a large network of users (e.g. friends of friends on Facebook, which for many is made up of thousands of users).

focusing on public comments on online articles. The results strongly indicate that expectation around the use of such comments is lagging behind use trends. For example, fewer than half of study participants expected their public content to be accessible via "social search", despite the existence of social-based search results in Google [14] and Bing [9].

In addition, we highlight two user experience challenges in implementing content publicity going forward. Service providers should increase visibility into content publicity, and second, this publicity should be clearly associated with the privacy settings that are under the user's control, so that the user makes a more informed choice. The former is important because personalization may create variation in the way users appear to others; aspects of a user's online activities may be emphasized or de-emphasized based on the recipient's own activities and interests. For example, a friend who is an active online gamer may see gaming activities highlighted even if those activities represent small portions of their friends' time and are not included at all in the public self-representations that their friends control (e.g. blogs, web pages). In addition, expectations are equally important because if they are not compatible with current practice then users might choose privacy settings that do not reflect their preferences for content visibility.

2 Related Work

Others have noticed the technology-enabled trend toward publicity (e.g. [25, 27, 31]. We offer evidence that user expectations are not keeping up with this trend in Section 4.

In [19], evidence of user experience degradation as a result of this trend is provided through observations of "publicly private" behavior on YouTube. Such behavior can only exist when content is public but not publicized. We discuss directions for remedying this situation in Section 3.

Finally, we note that there are many technological drivers of this trend including people search (e.g. [17, 22, 18]), social search (e.g. [9, 14]) and personalization of web sites (e.g. [8]).

3 Transparency and Personalization

When personalized content is pushed to users through their online activities (e.g. search, news feeds, etc.) it becomes more difficult for a user to gauge how they are perceived by others. Prior to personalization, a query by name through a search engine provided a reasonable gauge of the links most commonly associated with a given person, but with personalized search it is less

broadly predictive. In particular, any piece of content associated with a person may be promoted in significance if through personalization it is commonly pushed to users with related interests or attributes.

Users are clearly struggling with understanding and controlling content visibility. For example, on Twitter (where profiles and posts are public by default) questions about transparency figure prominently in the FAQ (e.g. "How do I know who is following me?", "Who reads my updates?") [30]. Similarly, questions about profile visibility and privacy are popular on LinkedIn [16] and Facebook [6]. Transparency questions are also prominent in web searches. For example, the searches "who can see posts" and "hide status updates" surged in 2009 and 2010, respectively, with continued high volume into 2011, according to Google Insights for Search [13].

Reputation monitoring companies (e.g. [24]) are attempting to address this problem in that they identify online information about a user of which the user may be unaware. However they search for user content broadly, and do not attempt to gauge the perspective a particular person may have of an individual. In addition, the "walled gardens" of many social networks make such a "friend perspective" quite difficult to achieve. Some of the more promising directions for a user-friendly approach to friend perspectives include the efforts toward open APIs for social services (e.g. [21]), tools that leverage developer APIs for transparency [12], and privacy settings that detail what information a particular friend can view about a user through a particular service [32]. However, privacy settings are notoriously difficult for users to configure (e.g. [20]), and such tools and APIs have not yet achieved a broad scope.

4 User Expectations Around Public content

While many users struggle to understand the use of their content, many others appear to be largely unaware that content may be used outside the context in which it is given. As evidence of this we discuss results from a user study of expectations around the use of public content.

The study was administered in the form of an online survey (details below) and does not exactly replicate the online services that motivate it, in addition, it is well known that asking users to self-report their behaviors/reactions can produce unreliable estimates [15, 2, 23]. For these reasons the percentages presented below should not be taken as hard predictors of user behavior, but rather the *overall* low magnitude of the reported expectations is strong evidence that user expectations are not compatible with current content use practice. We also note that care was taken in the survey to avoid inflamma-

tory text that might bias users; indeed, the term "privacy" does not appear in the survey, although clearly the survey has a strong privacy motivation.

To gather data points on user expectations around the use of their public content we conducted survey studies with 200 users in the US, 100 in Germany and 80 in Italy. The users were paid to take part in our study and come from a broad pool of testers, the majority of whom have college degrees and are within 24-45 years of age. Only slightly more than half of the pool is male. We do not have demographic information for the specic users who completed our studies. All studies were completed online with no direct interaction between the users and the authors of this paper.

Each group of users was shown the title, snippet and url of an article in the language of their country. They were asked to read the article and answer 9 questions (all in English) about their interest in the article and the topic of the article, their interest in sharing the article, their interest in posting public comments about the article, and their expected and desired uses of such public comments. They were also asked about their historical frequency of sharing online content in order to detect any differences based on sharing habits. We asked about posting comments in the context of an actual article in order to make the setting more realistic, and to hopefully increase the accuracy of the answers. In addition, doing so allows the detection of any preference differences correlated with posting willingness. For completeness we include the questions in more detail below, note that questions 7(b) and 7(d) are compatible with trends toward social content on web sites (e.g. [8]) and social search (e.g. [9, 14]):[2]

1. How interesting is this particular article to you? [5 answer options]

2. Would you agree with the following statement? "This article represents an interest of mine and I would like to spend time looking at more articles like this." [5 answer options]

3. How often do you share content and/or recommendations with your contacts? That is, do you email, microblog/blog, or share links (urls) via a social network (e.g. Facebook): [5 answer options]

4. Suppose that next to this article there was a "share" button. If you press this button, you will be asked to select one or more of your friends from a list, and they will receive a link to the article. Would you take the time to do this for this article?

5. Suppose articles like the one you just read include comments from other readers when they appear in

Google search results. That is, just as reader comments appear on publisher sites like the New York Times online, snippets of these same comments (each expandable to the complete comment) appear with the article in search results. These comments can be short indications of interest in the article (e.g. as with the Facebook "like" button) or more involved text. How likely are you to enter some sort of public comment for this article? [5 answer options]

6. Please expand on your response to question 5. What factors influenced your decision? [Text box for answers]

7. When you post a public comment on an article, which of the following can happen? Please select all that you believe are likely.

   (a) My search results will be personalized based on articles on which I've commented.

   (b) My friends will receive emails with links to the articles on which I've commented.

   (c) When my friends see this article in search results they will see I have commented on it.

   (d) Anyone who sees this article in search results will also see that I commented on it.

   (e) Anyone who sees this article in search results will see a count of all the comments, including my comment. No names will be given.

   (f) Anyone who searches for me and finds my online profile will see that I have commented on this article.

   (g) None of the above.

8. When you post a public comment on an article, which of the following would you like to happen? Please select all that apply.

   (a) My search results will be personalized based on articles on which I've commented.

   (b) My friends will receive emails with links to the articles on which I've commented.

   (c) When my friends see this article in search results they will see I have commented on it.

   (d) Anyone who sees this article in search results will also see that I commented on it.

   (e) Anyone who sees this article in search results will see a count of all the comments, including my comment. No names will be given.

   (f) Anyone who searches for me and finds my online profile will see that I have commented on this article.

---

[2]Wording is changed in the answers to questions 6 and 7 for anonymization.
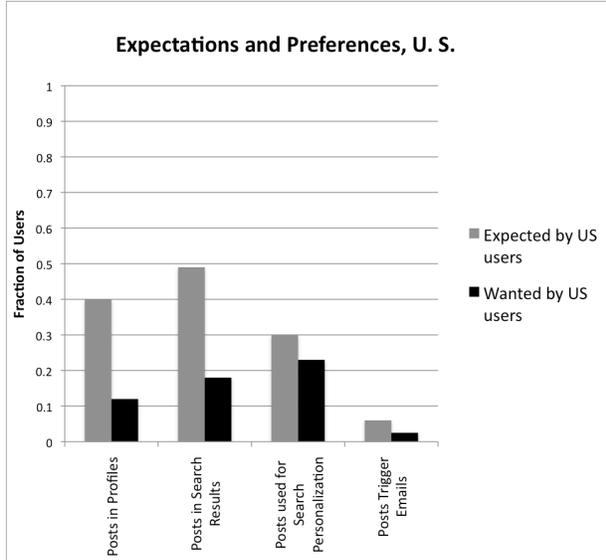
Figure 1: Less than half of the US survey participants expected their content to be publicized in the ways suggested by the survey (question 7). The content also indicates that the suggested uses are not desired, with less then 20% reporting they want their content publicized in the suggested ways (question 8). The x-axis labels map to survey questions as follows: "in Profiles" is option (f), "in Search Results" is (e), "for Search Personalization" is (c) and "Trigger Emails" is (b).

    (g) None of the above.

9. Please explain your answer to question 8. What do you like and what don't you like about linking article comments with articles? [Text box for answers]

RESULTS. We first limit our discussion to the results for US users for clarity of exposition and because they are representative of what we found overall.

The survey results do not indicate that users anticipate the publicizing of public content. On the contrary, less then half of the users anticipated their actions would be publicized in the ways suggested by the survey. Furthermore the survey participants strongly indicated they want to have control over how their content was publicized, with most users stating they would not want their public content publicized via emails and public profiles.

Note that although social search had already been deployed with respect the Facebook Like button [9] at the time of the study, it was not mentioned by any of the users in the study indicating low awareness of this feature. In addition, user expectations were quite low for the publicizing of public comments, a practice that is growing in popularity [5]. We see this both in the low expectations around posts appearing in profiles (the default on Twitter) and in responses to question 7(b), which asks the likelihood that a public comment on an article will trigger an email to friends.

Overall, the users did not see much value to the suggested content uses (question 8). A notable exception to this is the aggregation of public comments into a single count; this information was desired across all countries and was the most popular option in each country by a wide margin.

While value judgments on these uses were often low, it is important to note that users who reported being likely to comment on the shown article were the most positive about the uses, with differences that are statistically significant when compared to the responses of users who reported they were not inclined to post. For example 28% of those who reported being inclined to post valued search personalization based on their public comments versus 18% of those who reported not being inclined to post (p-value = .026) and 40% of the reported posters wanted friends to see their posts in search versus 24% of the reported non-posters (p-value = .003). These differences suggest that expectations around the use of public content are somewhat elastic based on context and perceived value.

The one use valued by the non-posters was the aggregate post count, perhaps because it is a privacy-aware popularity measure. 5% of the reported posters wanted aggregate post counts in search versus 39% of the reported non-posters (p-value = .02).

Results for users in Germany and Italy were similar to the US results overall. Most of the statistically significant differences across countries are shown in Figures 2 and 3. Figure 2 shows the content use options seen as most likely and Figure 3 shows the most preferred use options, with any statistical significance indicated. Note that even when the differences are significant, all the expectations are below 50% and the percentage who desire the use is below 25% in each country. In all the figures the vertical axis scale is $0-1$ to make it easy to gauge the overall fraction of users with a given response by visual inspection.

5 Summary

We've provided quantitative evidence that user expectations lag behind the trends in content use for personalization. Such an expectations mismatch increases the chance of privacy problems as users may configure privacy settings expecting incorrect outcomes. User outcry around sites like Spokeo [29] is one such example.

To remedy this situation we argue that publicity should continue but with increased transparency into content use. Work has begun in that direction with open API efforts and fine-grained privacy settings. However, user
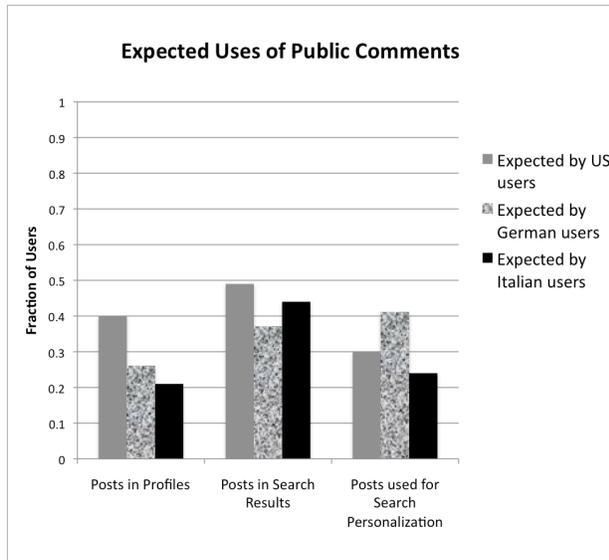
Figure 2: The content use options seen as most likely (question 7). Differences in the profiles category are statistically significant (p-value $< .05$) between the US and Germany and between the US and Italy. In search results, the difference between the US and German is *weakly* significant (p-value$= .055$). We also note that in search personalization, the difference between the US and Germany is *weakly* significant (p-value$= .058$).

difficulty in configuring such settings and current limits in their scope across services demonstrates a more usable solution is still needed.

The views expressed in this paper are those of the authors alone and do not in any way represent those of Google.

6 Acknowledgments

References

[1] M. Ackerman, L. Cranor and J. Reagle. Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences. in Proceedings of EC99 (Denver CO, November 1999), ACM Press, 1-8.

[2] J. Barabas and J. Jerit. Are Survey Experiments Externally Valid? American Political Science Review 104 (May): 226-42. 2010.
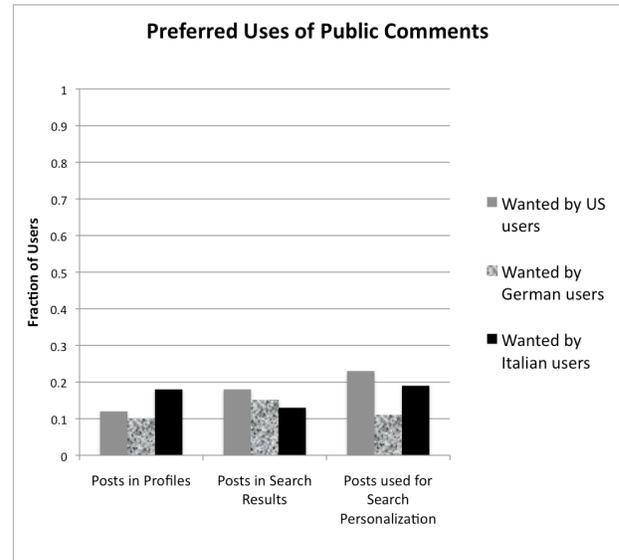

Figure 3: The most popular content use options (question 8). In search personalization the differences are significant between the US and Germany and between Italy and Germany.

[3] M. Benisch, P. Kelley, N. Sadeh and L. Cranor. Capturing location-privacy preferences: quantifying accuracy and user-burden tradeoffs. Personal and Ubiquitous Computing, December 2010.

[4] A. Das, M. Datar, A. Garg and S. Rajaram. Google News personalization: scalable online collaborative filtering. WWW 2007, Industrial Practice and Experience Track.

[5] Facebook Developers, Comments. https://developers.facebook.com/docs/reference/plugins/comments/

[6] Facebook FAQ. https://www.facebook.com/help/faq/

[7] J. Constine. Facebook launches new feature asking users how often they want to discover new games. InsideSocialGames. com, April 20th, 2011. http://www.insidesocialgames.com/2011/04/20/facebook-launches-new-feature-asking-users-how-often-they-want-to-discover-new-games/

[8] Facebook instant personalization sites. https://www.facebook.com/instantpersonalization/

[9] A. Ha. Bing and Facebook try to crack social search. VentureBeat.com, October 13,2010. http://venturebeat.com/2010/10/13/bing-facebook-social-search/

[10] S. Henig. The Tale of Dog Poop Girl Is Not So Funny After All. July 7, 2005. Columbia Journalism Review.

[11] C. L. Hovland. Reconciling Conflicting Results Derived From Experimental and Survey Studies of Attitude Change. American Psychologist, 14: 8-17. 1959.

[12] M. Ingram. Want to Know What Facebook Is Saying About You? Try This Tool. Gigaom, April 27, 2010. http://gigaom.com/2010/04/27/want-to-know-what-to-know-what-facebook-is-saying-about-you-try-this-tool/

[13] Google Insights for Search. http://www.google.com/insights/search

[14] Google Social Search help page. http://www.google.com/support/websearch/bin/answer.py?answer=165228

[15] C. L. Hovland. Reconciling Conflicting Results Derived From Experimental and Survey Studies of Attitude Change. American Psychologist, 14: 8-17. 1959.

[16] LinkedIn Help Center. https://help.linkedin.com/

[17] 123People. http://www.123people.com/

[18] 411. http://www.411.com/person

[19] P. Lange. Publicly private and privately public: social networking on YouTube. Journal of Computer-Mediated Communication 13(2008), pp 361-380.

[20] M. Madejski, M. Johnson and S. Bellovin. The failure of online social network privacy settings. Columbia Technical Report, CUCS-010-11. https://mice.cs.columbia.edu/getTechreport.php?techreportID=1459

[21] Open Social. https://sites.google.com/a/opensocial.org/opensocial/Home

[22] PeopleFinders. http://www.peoplefinders.com/

[23] M. Prior. The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure. Public Opinion Quarterly, 73 (1): 130-143. 2009.

[24] Reputation.com. http://www.reputation.com/

[25] Scientific American, Special Report. Technology's Toll on Privacy and Security. August 18, 2008.

[26] N. Sadeh, J. Hong, L. Cranor, I. Fette, P. Kelley, M. Prabaker and J. Rao. Understanding and capturing people's privacy policies in a mobile social networking application. Personal and Ubiquitous Computing, August 2009.

[27] N. Singer. Technology Outpaces Privacy (Yet Again). The New York Times, December 11, 2010.

[28] Spokeo. http://www.spokeo.com

[29] Spokeo website raises privacy concerns. Fox59 News, March 31, 2010. http://www.fox59.com/news/wxin-spokeo-website-privacy-concerns-033110,0,2433092.story

[30] Twitter FAQ. http://support.twitter.com/entries/13920-frequently-asked-questions

[31] J. Weintraub and K. Kumar. Public and private in thought and practice. Chicago and London: University of Chicago Press, 1997.

[32] Windows Live Messenger privacy settings. http://explore.live.com/windows-live-messenger-what-can-others-see-using?os=other