

Social Tagging: Implications from Studying User Behavior and Institutional Practice

Õne Mets¹  and Jaagup Kippar²

¹ University of Milano-Bicocca, Milan, Italy
o.mets@campus.unimib.it

² Tallinn University, Tallinn, Estonia
jaagup.kippar@tlu.ee

Abstract. This paper aims to describe users' tagging behavior in catalogues and in Flickr. Six platforms of two institutions and one consortium are analyzed: the main catalogue Discovery and Flickr page of the National Archives of the United Kingdom, the main catalogue Explore, the catalogue Archives and Manuscripts and Flickr page of the British Library, and the consortial search engine of the pan-European eBooks on Demand Library Network. The results of the document and user data analysis point to differences between archival and library collections, between catalogues and Flickr, illustrate the impact of different authorization and procedural rules, and confirm previous studies as regards to the small size of the active user group. Based on the data analysis, we offer eight recommendations for social tagging in libraries and archives concerning the issues of interface functionality and management, data collection, reflection of tags and maintaining the community.

Keywords: Crowdsourcing · Social tagging · Folksonomies · Catalogues · Flickr

1 Introduction

Metadata is considered a key feature to discoverability of the collections of cultural institutions [1, 2]. Traditional indexing techniques are costly and labor-intensive and may not provide the only or best way to meet user needs in online resource discovery [3–5]. Yet the information that needs to be described is diverse and voluminous. Numerous digitization projects in libraries [6] reveal undescribed aspects of the content, and archives may lack of item-specific information, while organizing their records by collections [7]. Then crowdsourcing is applied in the form of social tagging.

Crowdsourcing projects have become popular in the cultural heritage sector. Projects can be characterized by making collections available in smaller sets to achieve systematically sub-goals, progress of task completion is monitored and communicated, volunteers are motivated and sometimes specialized skills, knowledge or equipment is required [8]. The conclusions of Transcribe Bentham project illustrate the overall picture: majority of work is done by minority of users; volunteers have interest in the subject, crowdsourcing or the technology and sense of altruism; lack of time and issues with technology might limit participation whereas media attention increases it;

the project resulted in increasing the digital literacy skills of participants, contribution to scholarship and widen access to the material, adjusting workflows, and exploiting investments for digitization, software development and staff salaries [9].

Examples of ongoing tagging activity without the project-type framework provide interesting findings, but in a single institutional context: e.g. most tagged objects in a catalogue of Powerhouse Museum in Sydney were not on public display [10]; 67000 tags by 2518 people were attributed in the first 10 months in the Flickr page of the Library of Congress - a pioneer in Flickr, whose collaboration led to development of the Flickr Commons [11]. But the (collaborative) projects or ongoing activities in a single institution report different aspects of the results, which are difficult to compare.

This paper takes a comparative insight into the action taken by volunteers on six platforms of two institutions and one consortium: the main catalogue Discovery of the National Archives of the United Kingdom (TNA), the main catalogue Explore and catalogue Archives and Manuscripts of the British Library (BL), the Flickr pages of both organizations, and EOD Search, the consortial catalogue of the pan-European eBooks on Demand (EOD) Library Network. All cases represent social tagging as ongoing linear activity unlike the crowdsourcing projects.

The research questions are: What characterizes the tagging behavior, if it is not part of a crowdsourcing project? What are its affecting factors, outcomes, and implications? It contributes to research on online user behavior and folksonomies, and has a consultancy value for cultural organizations. However, the use of social tags by wider user community requires access to different data and is not studied in this paper.

The next section introduces the applied methods. Based on document analysis the platforms are described and the results of user data analysis are presented in Sect. 3. The findings are discussed and recommendations are proposed to archives and libraries for further consideration in Sect. 4. The paper is concluded in Sect. 5 with an outlook to the future work, where the current findings will be embedded.

2 Methods

2.1 Document Analysis

First, the six platforms were described referring to the interfaces, help articles alongside on the websites or linked pages. Some information was received directly from the institutions with delivery of data or by special enquiry.

The document analysis looked at 14 parameters: type of the platform (catalogue, social network site or other), the collection available for tagging (records, textual or non-textual items), online access to the items (full, restricted, partial or no access), collection size (number of items), pre-existing metadata, existence of application programming interface (API), releasing collections by small sets for tagging, time of launching social tagging, authorization of taggers (procedures of registration and sign in), publishing of social tags (immediate, verified), representation of tags to view or browse, procedure for deletion of tags, instructions to tag, syntax (separators of tags).

2.2 User Data Analysis

The acquired parameters for the user data were: tags, user IDs (anonymous for catalogues), item IDs, time of tag attribution (if recorded). The datasets with social tagging information in catalogues of BL, TNA, and EOD were composed by the respective institutions on request and delivered as separate CSV files. The dataset for BL Flickr account was composed earlier by the institution, and delivered as TSV file. The data for TNA Flickr account was extracted by using the Flickr API.

The data were imported to R [12] for analysis. All in all 25 parameters were calculated, including total and unique tags and tagged items per person, total and unique tags and contributing users per item, users and items per tag, returns and tagging activity per person across catalogues (for BL), correlations between parameters. Calendar converter¹ was used for calculating periods of returns.

3 Results

3.1 Document Analysis

Overview of the catalogues. BL main catalogue Explore² searches around 70 million items (records for books, journals, newspapers, maps, articles, Sound Archive items, Web Archive links etc.), being the biggest dataset in the comparison of the six platforms. TNA main catalogue Discovery³ holds over 32 million descriptions of records held by TNA (available for tagging) and more than 2500 archives across UK. BL catalogue Archives and Manuscripts⁴ includes unpublished documents, prints, drawings etc., the number of records in the catalogue is unknown. The EOD Search⁵ is a multi-lingual consortial catalogue, which runs on open source platform VuFind and searches over 7 million records of public domain literature from 35 libraries in Europe. The records link to institutional repositories for free full-text or display a button to request digitization for a fee [13]. Other catalogues in this comparison provide mostly limited or restricted access to view items. All are traditional catalogues with pre-existing metadata. No APIs are available for users.

Social tagging settings in the catalogues. Social tags were enabled first in Explore in November 2008, followed by EOD Search in the beginning of 2011, Archives and Manuscripts in January 2012, and Discovery in October 2012. Yet TNA may also be called a pioneer in this comparison due to launching a wiki site Your Archives⁶ in April 2007. A button was placed on the Document Details page of the catalogue taking to Your Archives to see if there was any additional information; otherwise, it created a

¹ <https://www.timeanddate.com/date/duration.html> .

² <http://explore.bl.uk> .

³ <http://discovery.nationalarchives.gov.uk> .

⁴ <http://searcharchives.bl.uk> .

⁵ <https://search.books2ebooks.eu> .

⁶ <http://yourarchives.nationalarchives.gov.uk> .

special page inviting the user to add content [14]. By 2012 the functionality was developed for Discovery, social tags were imported and the wiki was closed.

In the BL catalogues tagging requires sign in, which is only available to registered readers (registration can be completed in person at the Library⁷) and registered document supply customers (frequent users of the service, purchasing over 100 documents a year⁸). In Discovery anyone can register online, providing Reader's ticket number is optional. The EOD Search also enables anyone to register online.

Instructions about tagging are given briefly from each record's page in the BL catalogues. More detailed information is available from the opening page behind two clicks under 'Help articles'. The same information can be found behind the tab 'Tags', which is visible throughout navigation. Comma is required to separate multiple tags. In Discovery the record's page offers a link to sign in to add a tag, but detailed information about tagging is only available from the opening page. Another link after that page gives short tips about useful and appropriate tags, including instruction: "Simply enter a tag and click 'submit'. You can add as many tags as you like"⁹. In the EOD Search there is a note on a field in the record: "No Tags, Be the first to tag this record!". The only instruction in EOD Search for tagging appears after clicking 'Add Tag' button: "Spaces will separate tags. Use quotes for multi-word tags".

In all cases social tags are published immediately without verification, mostly next to the record, in EOD Search in a field in record. Tags can be deleted by the users, who attributed them and by the institutions. In the BL catalogues all tags can be browsed by most recent and most given. Logged in users can select to view only their own tags. In addition to tags BL enables to add notes, which are not indexed and not searchable, but moderated [15, 16]. TNA enables users to flag inaccurate tags, which are then checked by the staff. A spam and profanity filter to manage spam words is also in use. All tags can be browsed alphabetically, by most given and most recent.

Flickr. Both BL and TNA use Flickr to expose their selected collections of images. The BL Flickr account¹⁰ was established in August 2007 for corporate promotion. In December 2013 the BL Labs project added over 1 million undescribed images cropped from 65 000 volumes of digitized works from 17th to 19th century [17]. The experiment was meant for anyone to use, remix and repurpose and to spread new ways to navigate and display the content; and to stimulate the research concerning the materials [18]. First offered to Wikimedia, but rejected because of the lack of metadata, Flickr was chosen next because of tagging option, API existence, and attributing a unique URL for every image. BL imported to Flickr the metadata of the books, where the images came from, but there was no metadata about the images. Additionally, geotags are imported for maps from the BL crowdsourcing platform Georeferencer [19]. TNA joined Flickr in October 2008 and started to expose their thematic image collections since the beginning of 2011 "to give a flavour of their massive holdings"¹¹.

⁷ <http://www.bl.uk/help/how-to-get-a-reader-pass> .

⁸ <http://www.bl.uk/reshelp/atyourdesk/docsupply/help/register/regularcustomers/index.html> .

⁹ <http://discovery.nationalarchives.gov.uk/tags/index/howtotag> .

¹⁰ <http://www.flickr.com/people/britishlibrary> .

¹¹ <http://www.flickr.com/photos/nationalarchives> .

Anyone can sign up as a Flickr user and tag the images. Tags are displayed alongside the images as is the link ‘Add tags’. The tags added by Flickr robots are visible together with community tags, but distinguished by their white background. Next to ‘Tags’ under ‘?’ is a short description about tags and a link to some more information, including the instruction “Separate single word tags with spaces and add phrases in quotes”. Users can remove both tags they create and ones Flickr has added for them¹². The Flickr API¹³ enables anyone to write a program to present public Flickr data (photos, video, tags, profiles or groups) in different ways, and make their applications available to other users.

3.2 User Data Analysis

General overview. According to availability of the data, the period of observation varies as follows: 28 months (Dec 2013–Mar 2016) for BL Flickr page, 52 months (Oct 2012–Jan 2017) for Discovery, 60,5 months (Jan 2012–Feb 2017) for Archives and Manuscripts, 74 months (Jan 2011–Feb 2017) for EOD Search, 75 months (Jan 2011–March 2017) for TNA Flickr page, and 99 months (Nov 2008–Feb 2017) for Explore. Total numbers are presented for social tags and taggers on Fig. 1, because there is no significant distinction in the proportions compared to the results in average per month. The figure excludes 15% of total tags in Discovery (i.e. mostly tags attributed by users to Your Archives, then imported to Discovery by a single institutional user account), 43% of total tags in BL Flickr page (i.e. mostly tags attributed by users in Georeferencer, then imported to Flickr by a single account); 96% of total tags in TNA Flickr page (i.e. collection names etc. attributed as tags by TNA).

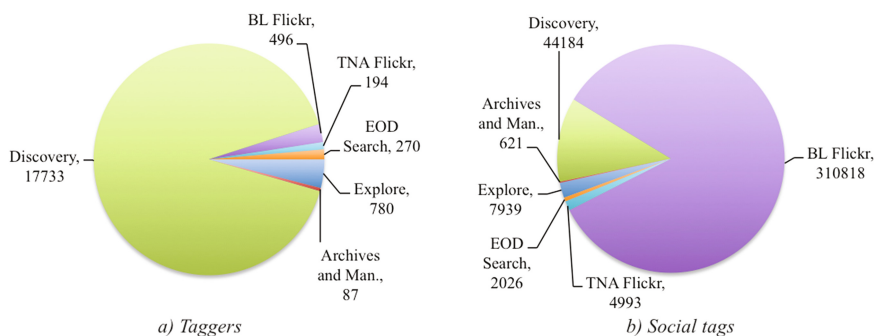


Fig. 1. Distribution of (a) engaged taggers, and (b) social tags attributed by them.

¹² <http://help.yahoo.com/kb/flickr/tag-keywords-flickr-sln7455.html> .

¹³ <https://www.flickr.com/services/apps/about> .

In all cases majority of users attribute up to 10 tags, and about 8 people form a group of top taggers (by attribution of total and unique tags or tagged items) (Fig. 2). If we exclude the tags by institutional accounts, the median value is 6 tags per person for BL Flickr page, 3 for TNA Flickr, 2 tags per person for Explore and EOD Search and 1 for others. BL Flickr page is also first by the sum of unique tags, but both Flickr pages have least unique tags out of total tags (TNA 5% and BL Flickr page 8%, Explore 26%, Archives and Manuscripts 28%, EOD Search 46%, Discovery 60%).

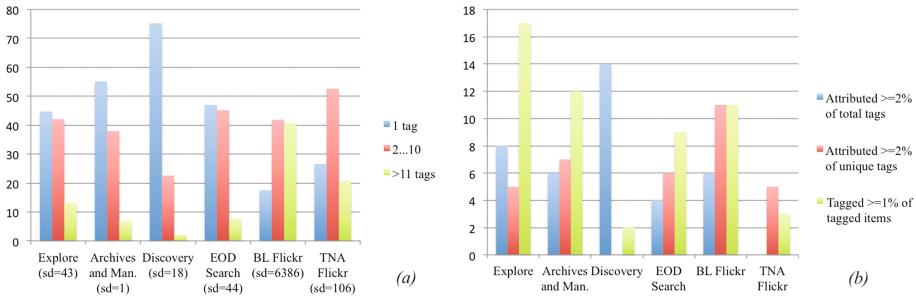


Fig. 2. (a) Number of tags per person (percentage of total taggers); (b) division of top taggers by total and unique tags and tagged items (absolute numbers).

Items. In average per month the most items were tagged in BL Flickr page (6290 items per month, 176 133 in total), less in Discovery (794 per month, 38 106 in total), in TNA Flickr page (209 per month, 15 679 in total), in Explore (56 per month, 5548 in total), in EOD Search (18 per month, 1360 in total) and in Archives and Manuscripts (9 per month, 528 in total). The tagged items gained mostly one tag per item in the catalogues (mean, median, mode < 1,5; sd < 2,2), more in Flickr (BL Flickr: mean = 3, sd = 68,5; TNA Flickr: mean = 7,2; median = 8; mode = 11, sd = 4,38). In all cases the items were tagged mostly by one person (mean < 1,42; median = 1; mode = 1; sd < 1,69). In Archives and Manuscripts no records were tagged by at least 2 different people. In Explore the maximum number of taggers per record was 4 (one occasion), in BL Flickr 5 people per image in maximum (18 occasions), in TNA Flickr page 8 people, but in Discovery there were 104 people per record in maximum (one occasion, followed by 100 people per record once, 80 people per record once etc.).

Correlation between tagged items and attribution of total tags is very strong in all cases ($r \geq 0,82$). Correlation between attribution of total and unique tags is strong in case of archival content (TNA Flickr $r = 0,98$, Discovery $r = 0,88$, Archives and Manuscripts $r = 0,72$), but remains moderate in other cases ($r \leq 0,46$). Similarly correlation between tagged items and attribution of unique tags is strong or medium for archival content ($0,8 \geq r \geq 0,57$), low for others ($r \leq 0,27$).

Time interval. Dates of tag attribution were analyzed for Explore, Archives and Manuscripts and EOD Search. In other cases the time parameter of tag attribution was not available. In Explore there were 956,5 total tags per year in average (sd = 529), in EOD Search 305 tags (sd = 315), and in Archives and Manuscripts 123 tags in average

per year (sd = 152). The number of engaged users and time interval for their returns in average per person are illustrated on Fig. 3. In all three cases users tagged mostly on one day, in average per person on $1,15 \leq \text{dates} \leq 1,58$ (maximum in EOD Search: 112 dates per user, sd = 6,79). The users, who returned to tag on a different date, are 165 people (21% of total taggers) in Explore; 13 people (4,81% of total taggers) in EOD Search, and only 11 people (1,4% of taggers) in Archives and Manuscripts.

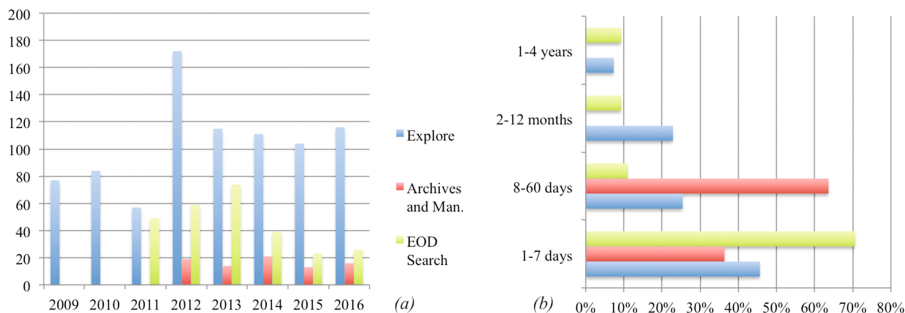


Fig. 3. Distribution of (a) engaged taggers by years, and (b) time interval for returns.

The period from first to last day of tag attribution varies as follows: 260 days in average per user (sd = 373, max. 3,7 years) in Explore; 252 days (sd = 489, max. 4,5 years) in EOD Search; 24 days (sd = 23, max. 2,3 months). The correlation between number of days from first to last day and days with tagging activity is stronger in Archives and Man. ($r = 0,68$) and Explore ($r = 0,54$), and weak in EOD Search ($r = 0,23$).

Correlation between the number of returns and attributed tags is modest (Explore $r = 0,46$, EOD Search $r = 0,36$, Archives and Manuscripts $r = 0,25$). For instance, the 18 top taggers (by total tags) in Explore divide into 3 people, who returned on >10 dates, 12 people on 2 to 9 dates, and 3 top taggers gave all their tags on one day. In Archives and Manuscripts one person added tags on 4 dates within 2 months. All other 10 of 11 users tagged on 2 dates and returned within the same or subsequent month, incl. the top contributor (attributed 55% of tags) on two subsequent dates.

Data for Explore and Archives and Manuscripts allows looking at tagging across catalogues. 16 people have attributed tags both to Explore and Archives and Manuscripts, i.e. 2% of taggers in Explore and 18% in Archives and Manuscripts. 2 people appear as top taggers in both catalogues.

Tags. The most attributed tags refer often to import of many tags at a time (e.g. ‘a level - korean war’ imported to Discovery from the previously run wiki site or ‘geo:continent = europe’ in BL Flickr page through its API), whereas tags attributed by most people differ in content, are sometimes private by their nature (e.g. ‘to read’) or point to the misuse of the syntax rules (e.g. not using quotes for phrases results in prepositions as tags, like in EOD Search) (Fig. 4).

Tags attributed most in total	Tags attributed by most people	Tags attributed most in total	Tags attributed by most people
Explore, BL		Flickr, BL	
silent cinema	dissertation	map	map
business directory	fantasy fiction	georefphase2	portrait
war of 1812	history	togeoref	church
telecommunications industry guide	read	rotate90	bird
this	to read	hasgeoref	london
anarchist newspapers & periodicals (english language)	biography	wp:bookspace=geography	castle
propaganda 2013 exhibition	check	geo:continent=europe	river
construction	music	coatofarms	dog
renewable energy industry guide	1	portrait	boat
Discovery, TNA		colorful	horse
a level - korean war	grandad	colourful	ship
a level - chartism	grandfather	people	woman
movcon	dad	lettert	cathedral
cycle	father	wp:bookspace=synopticindexusa	bridge
haiti list	ww1	rotate270	letter
EOD Search		geo:country=uk	man
incunable	1	geo:country=unitedkingdom	egypt
1	2	music	split
CS	of	Flickr, TNA	
officium	12	thenationalarchivesuk	woman
http://search.books2ebook	Geschichte	tna:departmentreference=co	horse
2	a	tna:seriesreference=co1069	railway
Band	des	tna:divisionreference=cod32	road
Ainsworth	in	africathroughalens	school
Hubertusburg	the	asiathroughalens	street
alphabets	von	asia	

Fig. 4. Tags attributed most in total and tags attributed by most people.

In Archives and Manuscripts only one tag ‘john’ was attributed by at least 2 people. If we exclude geotags by BL in Flickr, the list of top tags remains similar. And if we manipulate that dataset further by losing the computational parts of tag strings, more geographical locations appear as most attributed tags in total.

4 Discussion

The comparison includes two main catalogues of the memory institutions, their Flickr pages, a smaller archival catalogue of the Library and a consortial catalogue. The difference in the size of collections available for tagging does not seem to impact

directly the tagging activity. The difference in the type of collections may have increased the tags for BL Flickr collection of cropped images as it had no pre-existing metadata and users had freedom to add anything, even if the image discovery may have been more serendipitous compared to described collections.

The two institutions have different user authorization procedures in the catalogues. The low number of taggers in BL catalogues may be caused by the registration procedure, which requires personal presence or being a frequent customer of document delivery service, compared to Discovery where anyone can register online. Also Flickr offers free online registration, but engaged less users than the catalogues. The devoted subject communities may not be as used to use Flickr than catalogues.

Even though the datasets are different by the number of participants and tags, they result in similar number of top taggers by attribution of total tags (up to 14 people), unique tags (up to 11 people) and tagged items (2 to 17 people).

It is common in catalogues, that most users add only one tag. The Flickr users tend to add more tags per person, and there are more tags per item in Flickr than in catalogues - both trends possibly affected by the use of Flickr API.

It is evident in all cases that people, who add more tags tag also more items. Also given that only a few taggers per item we may conclude that users tend to describe items briefly and choose different items from each other. The phenomenal 104 users per a record in Discovery is an exception that may be explained by the biggest number of taggers in total - the higher the number of users the higher the probability of tagging the same object; or it may refer to users' mistake to attribute tags on a collection rather than a specific item level. It could have been explained by the collection size - the smaller the size of collection the higher the probability of people tagging the same item - unless Discovery exceeds the number of images in Flickr about 30 times.

The common feature for all three platforms with archival content is the attribution of unique tags compared to libraries' content, which can be described more by universal tags. It may refer to the perception of users to make a distinction of unique archival content compared to the published materials.

Comparison between most attributed tags and tags attributed by most people reveals interesting dichotomy. In four cases the most attributed tags refer more to applying computational techniques for tag attribution, mass import of tags from other sources or to the form of the tagged object, whereas tags attributed by most people tend to be more telling in content. E.g. Discovery suggests that the core interest of most taggers is genealogy. And even if tags like 'grandad' is a noise for others, it tells us to update the instructions with the most common examples. Additionally tags like 'to read', or 'alan's summer project' refer to marking up for individual need, suggesting to develop the functionality for private tags. Interestingly these examples occur in catalogues and not in Flickr, which has more liberal or noisy image than the controlled and verified one of the catalogues. It may also rise from the nature of the collections exposed in Flickr - selected images instead of records of materials.

The misuse of syntax for multiple tags or for multi-word tags may turn useful tags into noise. E.g. commas must be used in the BL catalogues, but if users follow the record and add 'Last Name, First Name' without quotes, it results in having a tag 'john' instead of full name. Similarly not using quotes for phrases in EOD Search resulted in having prepositions as tags. Flickr has the same rules, but these mistakes are not

common there. The crowdsourcing projects usually avoid this kind of mistakes by having separate text boxes for different descriptive data.

The data of three catalogues, where time factor was available, point to rather surprising finding that not only most users, but also most top taggers by total tags make their contributions within a short time-frame, less than 10 days and in some cases only once. The timescale does not imply that tagging has become more popular over time, rather it is unstable in terms of total tags and a bit steadier by participating users.

Additional content analysis of the social tags and analytics of the usage of the social tags could shed light to understand better the tagging under these conditions.

4.1 Considerations for Institutional Practice

Enable and instruct

- (1) Enabling tagging for everyone upon an online registration seems to have a positive impact on the number of contributors compared to ID verified users only. Institutional practices refer that volunteers come from different countries [8], so are the users of the catalogues, who are potential taggers. About 90% of the visits to the EOD search engine [13] and 30% of visits to Your Archives were from Google searches [14]. The description of the record may become more varied and more reliable, if tagged by more people.
- (2) Once the taggers come, they might not come through the first door, i.e. opening page of the catalogue, what was in one occasion the only place where to find instructions on tagging. EOD Search has experienced only 3% of the visits landing on the opening page, and majority of visits came directly to single records [13]. Help pages should be cross-linked.
- (3) Available and clear instructions proved to be vital especially for the use of separators. In some case comma was intuitively used instead of space, quotes were not used for phrases or users did not understand the instruction to insert one tag at a time. If user behavior suggests that the requirements for separators must be adjusted, it will likely not affect many users, who were used to different separators and will likely not return to tag in the future as suggested by the data.

Advance and reflect

- (4) Some tags even within top tags do not contribute to increasing discoverability, but are initiated from another motive - individual need. It is not clear, if users intentionally break the rule of 'making a useful contribution' or they have overseen the notion that all tags are made publicly available. Still users seem to need an alternative option to add some tags visible only to themselves, because these tags are not meaningful to other people ('to read' etc.).
- (5) As the digital skills of users improve (use of APIs, running software libraries, image recognition tools etc.), providing an API justifies the effort and significantly increases the amount of tags. It requires full availability of items, which in turn might reduce the risk of tagging based on assumptions when the item was not

fully seen. It may also lead us to an intriguing option to enable social (computational) tagging not in catalogues, but repositories.

- (6) Displaying tags, which are given by most people might be more telling than showing the most attributed tags by no matter how many people or technique. It is not the case when the amount of contributors is too low.

Monitor and maintain

- (7) Recording the time of tag attribution is important for monitoring the returns of the contributors. That data were available for 3 platforms of 6, all suggesting that not only majority, but even top taggers occur in short time.
- (8) If the goal is to keep the top taggers, their contribution should be detected quickly, and if seemed valuable, the dialogue should be started and maintained quickly before they leave.

5 Conclusion and Future Work

The variety of tags and especially the variety of tagged items, which becomes evident from this study, illustrates the importance of social tagging for the whole collection. The overall user activity confirmed the previous studies on the small proportion of active users. The current study also showed that driving tagging activity onto a social network site does not guarantee that the activity goes viral. The power of Flickr in this case lies in its API, which was used for mass tagging, but not in increasing engaged audiences, which could have been expected due to the social nature of the platform.

The take-aways for organizations suggest reviewing sign up procedures, making instructions clear and available, considering individual need for tagging, developing tools for computational tagging by users, defining “top tags” not only by their sum, but also by the number of people attributing them, monitoring the activity in time and cherishing the valuable contributors.

The results will be used for further research on determining the relations between social interaction and discoverability of resources. Next the research proceeds with social network analysis; thematic analysis of the interviews with BL and TNA for better knowledge of participatory practices [20, 21] and institutional activity systems according to Activity Theory [22]; and thematic analysis of the questionnaires to the users for better persona creation and understanding individual activity systems.

Acknowledgements. The author is grateful to the British Library and the National Archives of UK for the data, Baseer Baheer for his help with Flickr API, and Prof. David Lamas and Dr. Marco Gui for their feedback to this paper.

References

1. Higgins, S.: Digital curation: the emergence of a new discipline. *Int. J. Dig. Curation* **6**(2), 78–88 (2011)
2. Westbrook, R. N., Johnson, D., Carter, K., Lockwood, A.: Metadata clean sweep: a digital library audit project. *D-Lib Mag.* **18**(5–6) (2012). doi:[10.1045/may2012-westbrook](https://doi.org/10.1045/may2012-westbrook)
3. Matusiak, K.K.: Towards user-centered indexing in digital image collections. *OCLC Syst. Serv. Int. Dig. Libr. Perspect.* **22**(4), 283–298 (2006). doi:[10.1108/10650750610706998](https://doi.org/10.1108/10650750610706998)
4. Macgregor, G., McCulloch, E.: Collaborative tagging as a knowledge organisation and resource discovery tool. *Libr. Rev.* **55**(5), 291–300 (2006)
5. Daly, E.K., Ballantyne, N.: Ensuring the discoverability of digital images for social work education: an online tagging survey to test controlled vocabularies. *Webology* **6**(2), Article 69 (2009). <http://www.webology.org/2009/v6n2/a69.html>
6. Galloway, E., DellaCorte, C.: Increasing the discoverability of digital collections using Wikipedia: the pitt experience. *Pennsylvania Libr. Res. Pract.* **2**(1), 84–96 (2014). doi:[10.5195/palrap.2014.60](https://doi.org/10.5195/palrap.2014.60)
7. Szajewski, M.: Using Wikipedia to Enhance the Visibility of Digitized Archival Assets. *D-Lib Mag.* **19**(3/4) (2013). doi:[10.1045/march2013-szajewski](https://doi.org/10.1045/march2013-szajewski)
8. Ridge, M. (ed.): *Crowdsourcing Our Cultural Heritage*. Series: Digital Research in the Arts and Humanities. Routledge, London (2017)
9. Causer, T., Terras, M.: ‘Many hands make light work. many hands together make merry work’: transcribe bentham and crowdsourcing manuscript collections. In: Ridge, M. (ed.) *Crowdsourcing Our Cultural Heritage*. Series: Digital Research in the Arts and Humanities. Routledge, London (2017)
10. Chan, S.: Tagging and searching – serendipity and museum collection databases. In: Trant, J., Bearman, D. (eds.) *Museums and the Web 2007: Proceedings*, Toronto, Archives and Museum Informatics, 1 March 2007. <http://www.archimuse.com/mw2007/papers/chan/chan.html>. Accessed 24 Mar 2017
11. Oomen, J., Aroyo, L.: Crowdsourcing in the cultural heritage domain: opportunities and challenges. In: *Proceedings of the 5th International Conference on Communities and Technologies (C&T 2011)*, pp. 138–149. ACM, New York (2011). doi:[http://dx.doi.org/10.1145/2103354.2103373](https://doi.org/10.1145/2103354.2103373)
12. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2016). <https://www.R-project.org>
13. Mets, Ö., Gstrein, S., Gründhammer, V.: Increasing the visibility of library records via consortial search engine. In: *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 169–172. IEEE Press, Piscataway (2014). doi:[10.1109/JCDL.2014.6970164](https://doi.org/10.1109/JCDL.2014.6970164)
14. Grannum, G.: Harnessing user knowledge: the national archives’ your archives Wiki. In: Theimered, K. (ed.) *A Different Kind of Web: New Connections Between Archives and Our Users*. Society of American Archivists, Chicago (2011)
15. Explore the British Library. British Library (2014). <http://www.bl.uk/catalogues/search/pdf/tags.pdf>
16. Explore the British Library. British Library (2014). <http://www.bl.uk/catalogues/search/pdf/notes.pdf>
17. O’steen, B.: BL Flickr image dataset: User Submitted Tags (till March 2016). Figshare. <https://doi.org/10.6084/m9.figshare.3126481.v1>. Accessed 15 Oct 2016. https://figshare.com/articles/BL_Flickr_image_dataset_User_Submitted_Tags_til_March_2016_/3126481

18. O'steen, B.: A million first steps. Digital scholarship blog. <http://blogs.bl.uk/digital-scholarship/2013/12/a-million-first-steps.html>. Accessed 12 Dec 2013
19. Mahey, M.: Interview, Skype. 15 October 2016, 27 October 2016
20. Bonney, R., Ballard, H., Jordan, et al.: Public participation in scientific research: defining the field and assessing its potential for informal science education. A CAISE Inquiry Group Report, Washington, D.C.: Center for Advancement of Informal Science Education (CAISE) (2009). <http://www.birds.cornell.edu/citscitoolkit/publications/CAISE-PPSR-report-2009.pdf>. Accessed 24 Jan 2017
21. Simon, N.: The Participatory Museum. Museum 2.0. (2010)
22. Kaptelinin, V., Nardi, B.: Acting with Technology: Activity Theory and Interaction Design. The MIT Press, Cambridge (2006/2009)