Tagging, Pinging and Linking -

User Roles in Virtual Citizen Science Forums

## Abstract

This article investigates user roles in virtual citizen science projects through a case study of the Talk-forum of Shakespeare’s World. In order to address collective knowledge production, we study the use of hashtags, pinging and linking as a way of differentiating how researchers, moderators and different user groups use the forum according to their roles. We show how both volunteers and researchers have a much deeper interest in the texts that they transcribe and actively seek contextual information, shape new lines of inquiry and discover new phenomena. We conclude that the use of online forums in citizen science can play a crucial role for extending the knowledge production from academic research to a wider public interest, and also providing new knowledges beyond the assigned task of transcribing documents.

Keywords: citizen science; citizen humanities; discussion forum; Virtual Citizen Science

## Introduction

The rapid adoption of information- and communication technologies have created several new opportunities for researchers and volunteers to create knowledge collaboratively. Ranging from web-based solutions to mobile applications, several ways of discussing, disseminating and classifying information have been introduced, opening up new interfaces that barely existed only a decade ago. Contemporary online environments, from the social media giants to forums for niche interests, offer several technological functionalities that increase the prospects for communication among communities and topics, for example by tagging information, notifying users through “pings” and utilising hyperlinks to contextualise and connect instances of information. From the perspective of citizen science, understanding how researchers and volunteer contributors interact online and create knowledge together thus become key issues. Not only is this valid for discovering new configurations between different forms of expertise, but also for improving citizen science as a research method that can be used over a wide range of scientific fields of inquiry. In this article we will study a discussion forum of a citizen science project as a way of understanding the possibilities offered by new two-way forms of communication that extend the research process to include non-researchers. In recent years, several studies of citizen science have been concerned with identifying typologies and classifications of organisational elements, distinguishing citizen science from “traditional science” as it extends to open participation and data sharing (Franzoni & Sauermann 2014), or can be divided into categories, as Bonney et al. (2009) have argued, such as “contributory”, “collaborative” or “co-created”, depending on the level of involvement of citizens and whether projects are based on action (creating knowledge to influence civic agendas), conservation or investigation, launched on virtual platforms and if they have educational purposes or not (Wiggins & Crowston 2011). In the terminology of Wiggins and Crowston (2011), “virtual citizen science” (VCS) demands further investigation since this is a novel form of citizen science that is increasingly used, often with great expectations not only as a method for approaching large datasets, but also as a new way of communicating scientific knowledge and methodology to the general public.

However, most of these studies have departed from an organizational perspective. Less is known from a “bottom up” perspective, especially regarding the importance of discussion forums, which instead turns the attention towards what citizen scientists are *doing* as they participate in projects. However, a few notable studies have emerged, and the interest in what forms of knowledge citizen scientists generate on discussion forums has been sparked in recent years.

In a study by Tinati et al. (2015) the authors traced the development of the Zooniverse platform and derived several design claims that included the creation of the Zooniverse “Talk” discussion forums that tie in with the platform’s tasks and allow object-specific discussions. They note that experienced users can take over additional roles; for example, as moderators filtering questions and notifying the science team if the community cannot answer a question by themselves. In another study, Liberatore et al. (2018) analysed what they refer to as “communities of practice” in a Facebook group administered by researchers. The authors found that the users “used the group to share excitement, ideas, and knowledge about New Zealand garden birds [...]” (Liberatore et al. 2018: 11). Moreover, some studies have performed data-driven analyses of VCS discussion forums (e.g. Luczak-Roesch et al. 2014; Ponciano and Brasileiro 2014). For example Luczak-Roesch et al. found that VCS has a small number of users contributing a large percentage of the activity (Luczak-Roesch et al. 2014). Hedges and Dunn call such users “super-contributors” (Hedges and Dunn 2017), as a contrast to a the larger number of users who make fewer contributions in total, and often only one or two contributions each. Reed et al. investigated the number of minutes spent on classifying data in the Galaxy Zoo project, and used this distribution to form a stratified sample of users, which could then become the target for their survey on motivation (Reed et al. 2013). In a similar vein Ponciano and Brasiliero (2014) used four different metrics of user activity to distinguish “engagement profiles” among the users in the projects Galaxy Zoo and The Milky Way Project.

These data-driven approaches, in which user contributions are divided into segments with a quantitative approach has both benefits and drawbacks. While it identifies users that are devoted and spend a large amount of time in the project as well as giving a good measure of the contributions made by the so called “long tail” (ie a large number of users that only make a few contributions each), it is insensitive to the particular roles that have functional differences in a discussion environment. In our case, such roles are moderators and researchers. Their roles in a forum discussion will not be detected by quantitative measures alone, since they may show similar patterns as both super contributors and, on the other side of the spectrum, as casual users. Yet, their functional roles can be crucial for the dynamics and knowledge production in a VCS forum. For example, a researcher will bring in both a specific expertise, often valued as authoritative, while moderators have a duty to keep discussions on topic and resolve emerging issues and conflicts. None of these qualities are readily detected through dividing user contributions in segments with a quantitative approach.

However, there are other non-linear aspects of forum environments that cannot be quantified in a simplistic manner. With non-linear we mean that there are affordances and functionalities in VCS forums that have bearing on the characteristic structure on the forum and its users. Hashtags (#), introduced on multinational platforms such as Facebook, Twitter and Instagram have entered the realm of citizen science. Studies have shown the importance of hashtags as markers of content, symbols of community and “influencers” (Zhang, Zheng & Pang 2018) and, in terms of learning, reflecting users’ needs over the duration of hashtags (Veletsianos 2017). Patterns created by hashtags have been found to be both stable and replicated as well as allowing for minority opinions (Golder & Huberman 2006) as they often can be used without constraints, but also developing into tag clouds marginalizing groups and individuals (Sinclair & Cardew-Hall 2008) since they give preference to more commonly used hashtags. On Twitter, research has shown that scientific authority is likely to lead to virtual authority, and the “hash tagging habits” of researchers participating in scientific conferences have been found to mainly be directed at peer researchers (Letierce et al. 2010). Studies of museum curatorial contexts have shown that terminology of tagging by contributors differs from that of professional curators (Trant 2009, Trant, Bearman & Chun 2007), implying different user behaviours.

With regards to citizen science, Hedges and Dunn have pointed out that the organisation of knowledge in VCS has pointed out that collaborative hashtagging serves as an important aspect of organising co-productive activities (Hedges & Dunn 2017).

For example, on the Zooniverse project Shakespeare’s World there is a clear distinction between tasks for the volunteers (transcribing text from scanned images of individual pages) and the researchers who analyse the refined and compiled data. However, the discussion forum, called “Talk”, allows for a second, more open and collaborative form of knowledge generation where volunteers can discuss interesting findings amongst each other and with the project’s researchers. This form of knowledge generation has led to several discoveries made by volunteers on the Zooniverse platform (e.g. Tinati et al. 2015: 4072). Thus, it becomes a crucial issue to further investigate the impact of VCS forums with regards to co-creation of knowledge.

## Purpose and research questions

The purpose of this study is to analyse how VCS discussion forums are used, with particular focus on user roles in relation to co-creation of knowledge. We intend to fulfill this purpose by investigating user behaviour in the Shakespeare’s World discussion forum in relation to the conventions and technical features offered by the platform. This includes the use of hashtags (#), the use of pinging (notifying users about discussions, so-called @-messages), and the use of external resources, which in turn offers new ways of systematising knowledge and discussion on VCS forums. In order to meet this purpose, we will structure our analysis according to the following research questions:

1. How are user contributions distributed among different roles?
2. In what way are external knowledge resources (for example links (hypertext)) used to bring in external information?
3. How are hashtags used and distributed over time?
4. How are pings (@-messages) used to interact and generate knowledge?

## Method and material

Shakespeare’s World is a citizen science project hosted on the Zooniverse platform. In this project, participants are invited to transcribe historical documents written by Shakespeare’s contemporaries, allowing both researchers and volunteers to learn more about the historical context of Shakespeare’s work. Additionally, the project aims to record new words and word variants for the Oxford English Dictionary (OED).

The data collected consists of 11450 posts from the project’s online discussion forum “Talk”, covering the activity within the first two years of the project (2015-11-09 to 2017-11-21). The data was exported using the built-in administrator function of the Talk software as a .json file, which was subsequently converted into various formats for analysis, such as spreadsheet files and textual corpora for rapid and specialised searches. The data was analysed using Microsoft Excel®, the Python programming language including a range of software libraries, and finally, the data visualisation software Gephi (Bastian et al. 2009) (see Appendix 1).

### Scheme of analysis and methodological approach

The Zooniverse Talk forum has a number of technological affordances that shape the departure point of our investigation. They can be summarised as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Threads | #hashtags | @-messages | Hyperlinks |
| Function | Organise subject matter | Sort and classify data | User interaction and notification | Referencing information |
| Empirical indicators | Length of threads, thread initiators (users). | Extraction of hashtags, frequency of use, clusters of users and hashtags. | Extraction of @-messages, frequency of mentions, networks of users and hashtags. | Extraction of http://-links, mentions of external resources. |

**Table 1.** Overview of functionalities in *Talk* and the empirical indicators used in the present study.

These functionalities form the software basis for the Talk forum and we have used them as ways of instantiating our line of inquiry. However, we will study how these functions are used in practice rather than taking for granted that they are used as intended or programmed. We will proceed along similar lines as Hedges and Dunn note, that “the rise of social media, especially multinational platforms such as Facebook, Twitter, and Instagram, has introduced the hashtag into the crowd's daily consciousness and has had a significant effect on the dynamics of tagging.” (2017: 34). In other words, by extending this analysis to include pings, threads and linking practices, we have structured the quantitative parts of our analysis on the extraction of empirical indicators as outlined in Table 1 by quantifying the length of threads, the frequency of hashtags, the frequency of @-messages and the frequency of links to webpages. Moreover, we have created two other data structures, one that is temporal in which we study the development of a hashtag over time, and the second is of a network character, to connect on a user-to-user basis the pinging practice in various hashtags. These types of data are necessary in order to properly address our research questions, in particular RQ 3 and 4 require more than just frequency measures, but must also include timestamp data and in the latter case the creation of network structure data.

### Ethical considerations

The Talk forum user agreement (<https://www.zooniverse.org/privacy>) grants researchers to use information entered to the Zooniverse platform (which includes its Talk pages) for the advancement of knowledge. Data exported from the Talk forums contain no personal information (such as IP-addresses, cookies, e-mail addresses) other than the username selected by the users themselves. In this article, however, we mention usernames only in relation to researchers who have presented themselves on the Shakespeare’s world “About” page. All other users are only presented as “moderators”, “super-users” et cetera.

## Result and analysis

As a first step we extracted and counted the frequency of use of the forum functionalities. A total of 11450 posts were written by 388 individual users in 3460 threads across 11 subforums (“boards”). Worth noticing is that threads are often short. There are only 10 threads with >30 posts and the average thread is only 3.3 (median 2) posts per thread. 972 posts (8%) contain hyperlinks, 2692 posts (24%) contain hashtags, and 2483 posts (22%) contain the @-symbol used to *ping* (send a notification to) other users of the Shakespeare’s world Talk Forum[[1]](#footnote-2).

## RQ 1 - Distribution of roles

For our study, we distinguish three groups of users: super-users (users contributing more than 100 posts to the forum each), active users (more than 10 posts), and casual users (10 or fewer posts). We arrived at the threshold of 100 posts for super-users as this was the approximate cutoff value for those users that had created 80% of the total amount of posts on the forum as a whole. Additionally, moderators and researchers in the project were identified and considered as separate roles. Table 1 shows an overview of the roles and their quantitative contributions to the Shakespeare’s World Talk forum.

A challenge in defining roles is that forums are dynamic and change over time. Moderators sometimes quit and new ones are assigned the task, and in Shakespeare’s world there is even a case where one user became a Researcher (this event is however not in our current dataset). The exported data retrieved does not contain the metadata for which roles each user has. In order to reconstruct roles we asked one of the principal investigators (Victoria van Hyning) to give us the dates (sometimes approximate) of when the roles were changed or assigned. This way we were able differentiate posts made by moderators, researchers and super-contributors, also marking up posts with the current status in cases where such roles have changed. While researcher and moderator are roles that are assigned as special qualities, our notion of super-users, active users and casual users are quantitatively defined (see above).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **User role in forums** | **Criterion** | **Number of users** | **Number of posts** | **Percentage of posts\*** | **Posts per user** |
| super-user | >100 posts | 11 | 4226 | 37% | 384 |
| active user | >10 posts | 36 | 1173 | 10% | 33 |
| casual user | <=10 posts | 327 | 805 | 7% | 2 |
| moderator |  | 4 | 3114 | 27% | 779 |
| researcher |  | 10 | 2132 | 19% | 213 |
| **TOTAL** |  | **388** | **11450** | **100%** | **30** |

**Table 2**. Community roles in Shakespeare’s World Talk (N=11450). While the number of posts takes into account the changing roles over time, this is not the case for the Number of users. Here the moderator and super-user roles have changed over time, for example one super-user became moderator, and two moderators quit over the course of the current dataset. \* As pointed out previously we used the 80% cutoff level as the definition of a super-user. Here such contributions include a subset of moderators and researchers.

The super-users produce 37% of the total forum content, thus, the contribution of these 11 users is substantially greater than the total of 363 other volunteer users. The second-largest group is moderators, although it should be noted that the 27% of forum posts by moderators are split up between only two individual users. Additionally, ten researchers contribute with 19% of the forum posts; four of these researchers have contributed with more than 100 posts each. Finally, the total number of contributions by 36 active users and 327 casual users only account for 10% and 7% of the forum contents respectively.

These results indicate first and foremost that the majority of the forum is created by a small number of users (17 users have created 80% of the posts). Among this group of productive users, we find the two most active moderators, four researchers and 11 super-users. The distribution among these roles is, however, uneven. In total the super users have created 4226 posts among 11 users and the four moderators have written 3114 posts, with the most active moderator contributing 2414 posts alone. All in all this puts these two moderators at a much higher individual production in comparison to other roles.

### Thread initiation

When the Talk-software was implemented, users were given the option of starting threads based on subjects encountered in the classification tasks instead of adding threads to a linear forum structure. This style of use is reflected in the Shakespeare’s World forum, where 92% of the threads were initiated based on specific subjects. Another pattern in relation to thread initiation is visible in the user roles. Table 3 shows the number of threads started by each group of users. Compared to the overall activity of the forum, the researcher-role stands out clearly: While the researchers contributed to about 19% of the overall forum posts, only about 1% of the threads in the forum are initiated by researchers. For the other user groups the distribution is more even. Moreover, the large bulk of subjects being transferred from the transcriptions to the forum are initiated by super-users (41%) and moderators (29%), which reveals that these users are the ones driving the creation of new materials that are in the need or interest of being disseminated to the rest of the community. This way, it is possible to conclude that the researchers have a different role in knowledge formation. They rarely start new threads, but instead frequently respond to questions raised as the threads develop. This is also the case for moderators and super-users, who do both tasks.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Thread initiation** | | **Thread responses** | | **Threads initiated per person** | **Thread responses per person** |
| moderator | 1001 | 29% | 2113 | 26% | 250 | 528 |
| researcher | 46 | 1% | 2086 | 26% | 5 | 209 |
| super-user (>100 posts) | 1413 | 41% | 2813 | 35% | 128 | 256 |
| active user (>10 posts) | 551 | 16% | 622 | 8% | 15 | 17 |
| casual user (<=10 posts) | 449 | 13% | 356 | 4% | 1 | 1 |
| **Total** | **3460** | **100%** | **7990** | **99%** | **9** | **21** |

**Table 3:** Threads initiated by each user group (N=11450). Note the percentages for thread initiation are calculated in relation to the total sum of threads (N=3460) whereas the thread responses are calculated using the total forum post which are not thread starts, but instead are responses (N=7990).

This brings about another interesting finding. The casual users (sometimes referred to as the “long tail”) have started a substantial amount of threads (449) but very few when broken down on each individual (in total 327 users). A similar pattern is found among the active users. They both have approximately a 1 to 1 ratio between thread initiation and thread response, which indicates that they drop in and out fairly quickly. Researchers, on the other hand, have the opposite behaviour. Starting up only 46 threads but writing 2086 posts indicates that they act as expert advisors and make their contribution once the subject has been brought to attention as a new thread. Moderators and super-users instead bring out many new threads simultaneously as they write in the forum twice as much as initiating threads. This means that not only are they very active in bringing out new data for discussion, but also play an important and large role in discussing, disseminating and analysing the data. Without the moderators and super-users, there would barely be an active knowledge production on the forum.

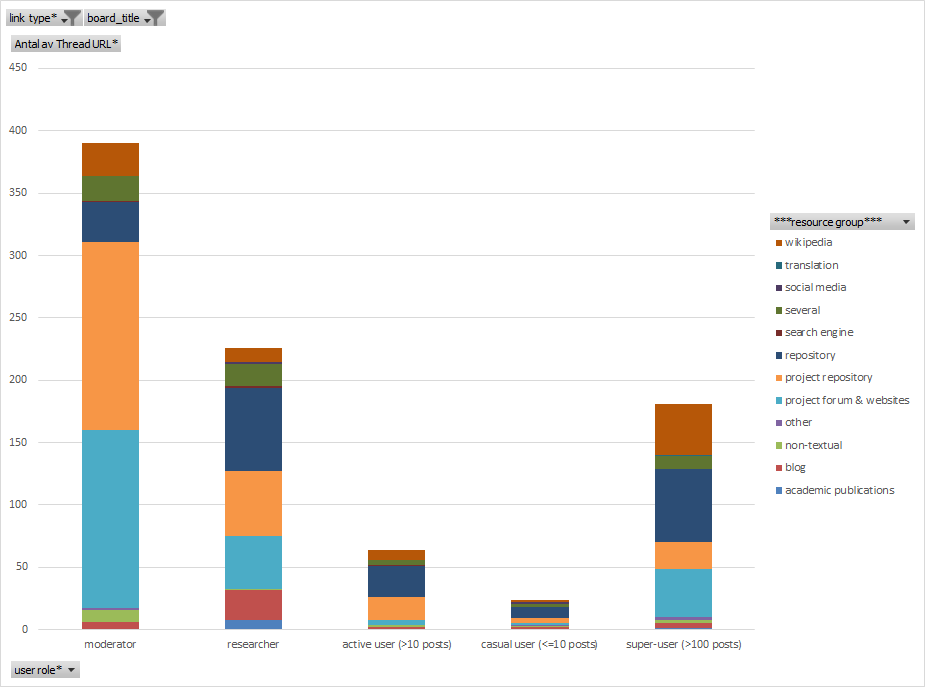
## RQ 2 - Use of external knowledge resources

Understanding letters written in the 17th century requires contextual information, especially for verifying word meanings, checking historical dates and facts, and to understand which historical figures are present in the texts. Such knowledge lies outside the task of transcribing the documents, however, as we will show, they are valued highly by the volunteers and researchers alike. As the web contains numerous sources of information that are external to the Shakespeare’s world forum, we have analysed the practice of bringing in external knowledge resources.

In this analysis we excluded the forum section “Help and Technical Issues” in order to better capture the knowledge practices that are directed to the research theme of Shakespeare’s world. Firstly, we extracted all URLs by searching for the ‘http’ prefix. Secondly we expanded our searches by including indirect links (such as “I used Google to find...”)(see Appendix 2). Then we coded them manually by the type of service used as an external resource, such as databases, archives or social media. The results are presented below in a two step sequence, where we present hyperlinks first and indirect links secondly.

### Hyperlinking

Out of 10605 forum posts 885 (8%) contain one or more hyperlinks, marked by the character string ‘http’. We investigated under which circumstances these were used, and also if the link behaviour differed between user groups.

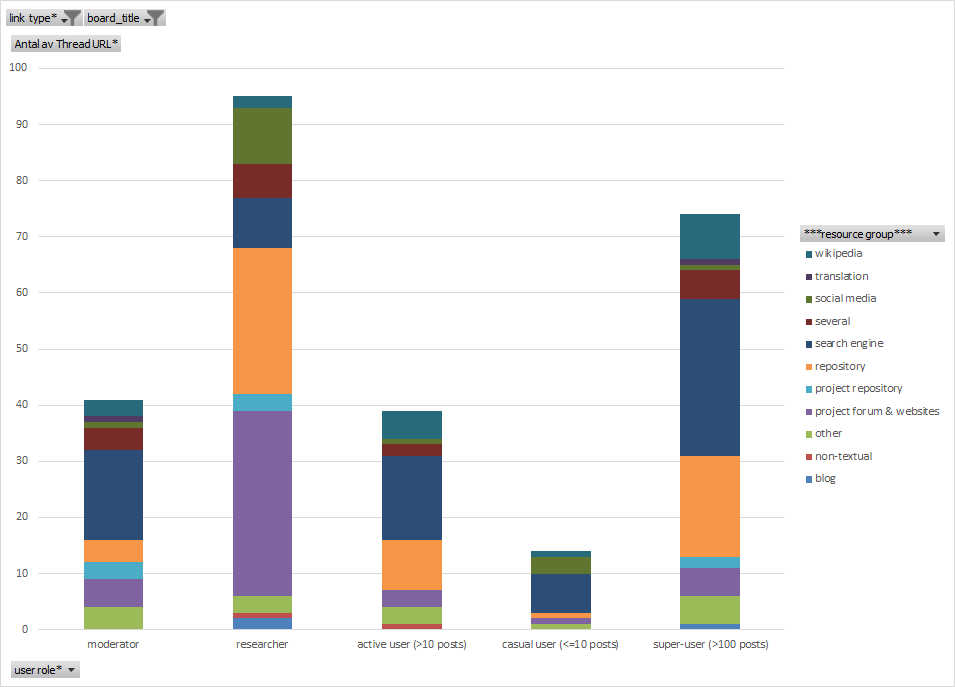


**Figure 1.** Use of internal and external hyperlinks containing the prefix string ‘http’ (N=885).

As shown in Figure 1, the different roles on the forum have different practices of bringing in information and linking to further sources. Moderators mainly link to various project repository documents, what is referred to a “subject” ie. a document that is up for transcription. Secondly they refer to the project forum and websites, linking together various threads of discussion. These two major linking patterns indicate that moderators are indeed fulfilling their assigned task in bringing together and moderating the forum by shaping more coherent discussions. Researchers, on the other hand, are quite diversified in their linking practices. While they also perform the roles of the moderators to a certain degree, as mentioned above, they frequently refer to the project forum and websites, comprised of instructions, blogs and other contextual information related to Shakespeare’s world and the Zooniverse platform. This way, they anchor the citizen science contributions within their project, and show how the knowledge generated becomes valued and fed back into research. Often they thank the contributors by updating the website or writing a blog post about a discovery or phenomenon in the transcribed documents. Researchers are less active in referring to other forum posts, while they more often than moderators refer to historical repositories and databases such as Wikisource, Project Gutenberg and Wellcome Library Documents, as well as dictionaries and thesauruses. They often mention several sources in the same forum post. This indicates a somewhat typical researcher style of writing, in which multiple sources to matters of fact are presented, and along the lines of knowledge synthesis, referring to an ongoing and cumulative research frontier. Sometimes this way of bringing in external resources are used by other roles once introduced as reliable sources by researchers. One such example is a link to commonly used symbols for measurement units in recipies (<http://www.textcreationpartnership.org/docs/dox/medical.html>), which was introduced by a researcher and then consequently used by super-contributors as a standard for interpreting such symbols. Furthermore researchers also link to blogs, which they often are the authors of themselves, and to scholarly articles. Finally, the super-contributors are quite diversified in their linking practices. We find a more frequent use of Wikipedia and the largest part of the links go to external repositories. This indicates that the super-contributors are wider in their quest for contextual information, going outside the project resources, often turning to free and open services such as Project Gutenberg, Wikipedia, and the Internet Archive.

### Indirect linking

When users refer to online sources and search engines without hyperlinking (ie. printing out the URL beginning with ‘http’) we call this indirect linking. It occurs more than 200 times in the forum material, and is often expressed in the style of “I used Google”, “I found on the Internet” or “I looked it up on Wikipedia”. As this occurred frequently, we deemed it necessary to be the object of further analysis. However, the extraction of such phrases is not as straightforward as with conventional hyperlinks. We created a set of search queries (Appendix 2) based on the first retrieval of hyperlinks. For example, if users linked [http://twitter.com](http://twitter.com/) or [http://](http://archive.org/)google.com, we generated search queries that would capture also indirect linking, in these cases “twitter, “googl\*” et cetera. This way we were able to exhaust most expressions, even though a limitation of this method is that we might miss on ways of linking expressed by phrases unknown to us. However, while informal linking is rather imprecise, it does indeed reveal a lot of interesting information about what forms of knowledge resources the users of the forum express as auxiliary resources that have a potential or actual capacity to bring in contextual information when transcribing and understanding the raw textual material.



**Figure 2.** Use of indirect links (N=263, see Appendix 2 for search terms).

The results of the informal links (Figure 2) show firstly that it is used by researchers the most, followed by the super-users. However, researchers tend to link indirectly back to the project forum and its websites (especially the Shakespeare’s World blog), to various repositories (especially the OED) and to social media while super-users as a contrast most often refer to search engines, a behaviour shared with moderators. Moreover, moderators, active users, casual users and super-users often refer to Wikipedia in their posts, while this is less common in the linking practices of researchers. This suggests that researchers are less willing to refer indirectly to free and open search engines and Wikipedia, while this appears to be favourable by all other users. Also, when breaking down the repositories, we find a stark contrast between researchers and super-users. While researchers point to the OED, the super-users more often refer to open repositories, such as the Dictionary of Scots Language (http://www.dsl.ac.uk/). We often find the expressions “on Google”, “the dictionary” or even “on the internet”, which are imprecise and sometimes impossible to know exactly what is referred to (which website, which dictionary etc.).

These indirect links are interesting to look at in a greater detail, since web searches can be used for almost any query, and appear to be utilised when the user is not sure where to look or how to exactly phrase the question. Notable examples among super-users and moderators include:

“The internet says this is Esztergom in Hungary. Its medieval Latin name was Strigonium and it has a history of battles against Turkey. ” (moderator)

“A quick Google tells us that the Beaumont family were indeed involved in coal mining so a foray into Welsh mining sounds reasonable - if unprofitable!” (moderator)

“I Googl'd the word ayenst and found that it occurs in the Canterbury tales. From the context, it seems to have the same meaning as "against"” (super-user)

“So. . . Anyone know what types of cows were around in Shakespeare's time? The "red" ones seem to have been singled out for this recipe implying that there were others.

I could Google. In fact I might. :)” (moderator)

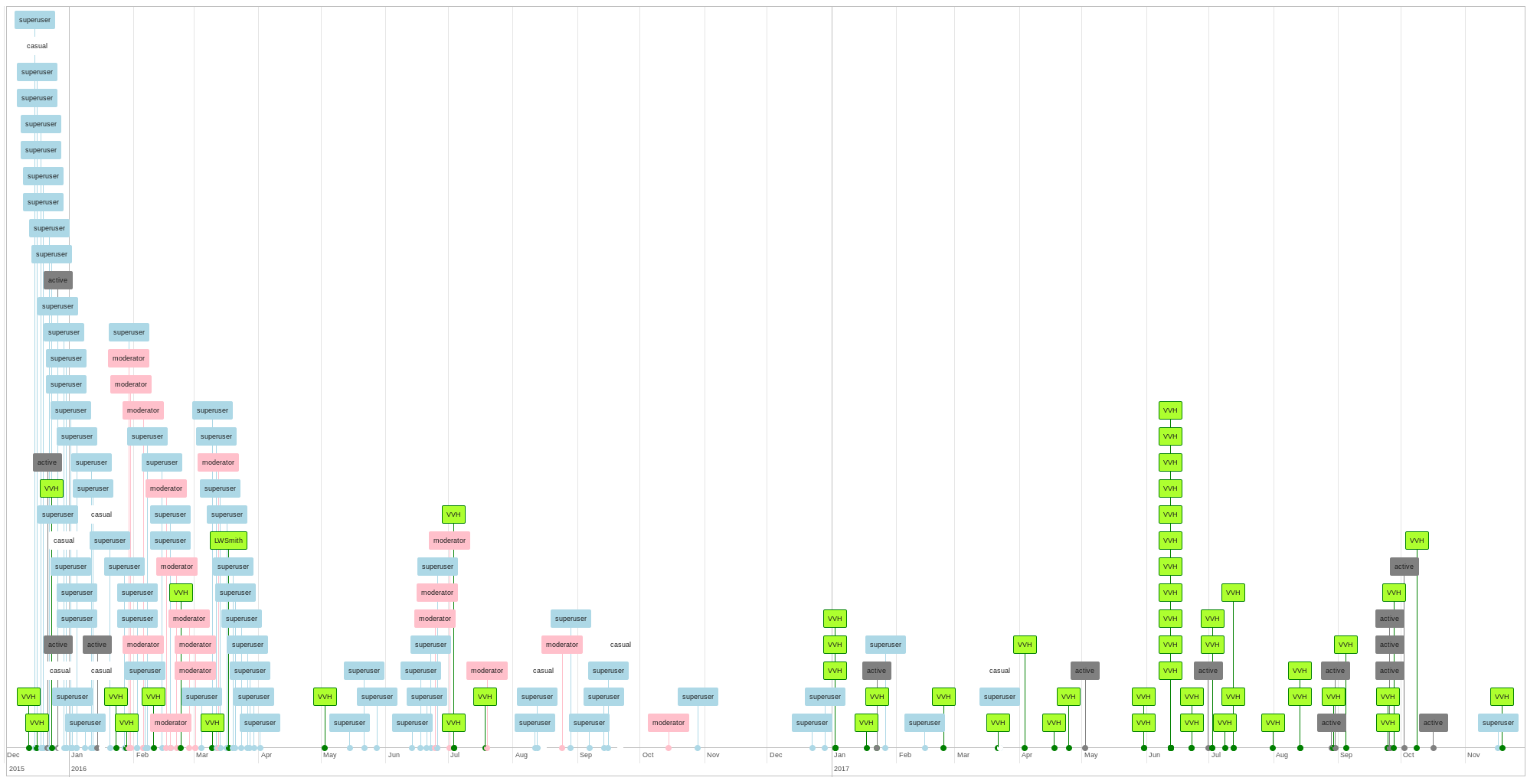
The overall practice suggests that searches are used to find contextual information that makes the transcribed text more meaningful. As in the examples above, users wish to find a historical place, a historical family or find the meaning of an old word which is not found in the average dictionary. This is also evident in the super-users’ frequent references to historical sources and documents which are brought in as points of reference.

## RQ 3 - Use of hashtags

Using hashtags is a common affordance in social media, for example Instagram, Facebook and Twitter, and has been implemented in the Shakespeare’s world forum. In this section we are interested in how hashtags are used to systematise knowledge, and if the practice differs between user groups.

### 1. Hashtags primarily used by researchers, moderators and super-users

The hashtags that are predominantly used by researchers, moderators and super-users are: #catholic, #OED, #paper, #womanwriter[[2]](#footnote-3). Out of these tags, #catholic, #OED, and #womanwriter were created as sub-forums from the beginning of the project and taken up by users as the project moved on. These all have in common that almost all tags are created by these categories of users. Often the super-users are the ones creating the most hashtags in quantitative terms.



**Figure 3:** Timeline for the use of the #catholic(s) hashtag. Green - Researcher, Pink - Moderator, Lightblue - Super-user, Grey - Active user, White - Casual user.

In Figure 3 we see the #catholic(s) hashtag as it is played out in our dataset over time. The first use is made by a researcher (Victoria van Hyning), but is almost instantly picked up by a handful of super-users, and a bit later by moderators. Most of the #catholic(s) post are made in the first six months of the project (this is also the case for the total data produced). However, in mid 2017 there is a spike in researcher contributions again, and when looking closer to these posts we see that the researcher is tagging up older threads by replying to users who have found interesting texts mentioning catholic writers.

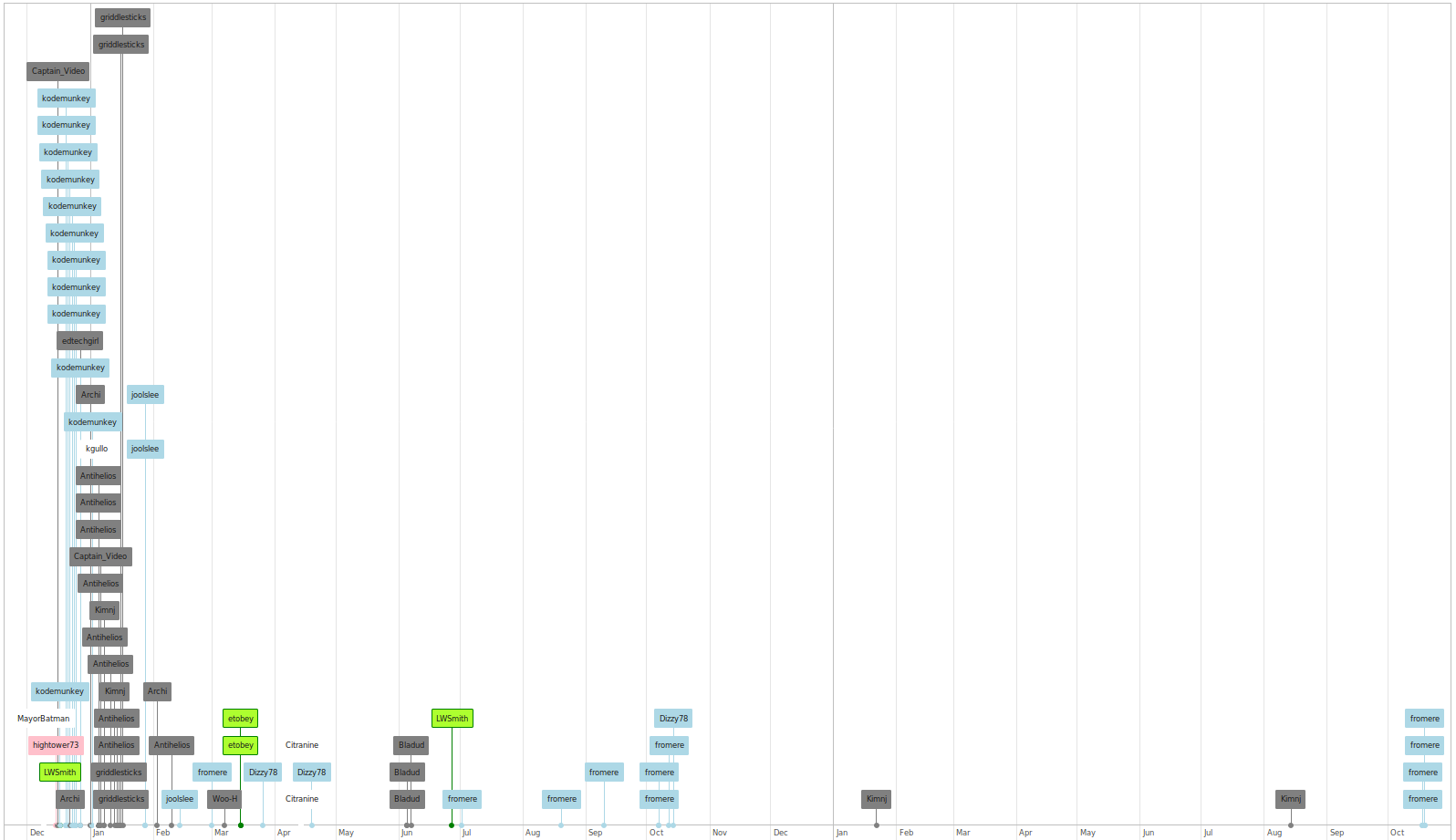
The #OED hashtag follows a different trajectory. It is heavily used in its first six months by researchers, who are soon aided by moderators and super-contributors. However, a year later it is once again picked up by another constellation of users, consisting mainly of active users, super-contributors and moderators. The #OED hashtag is of special importance because it is connected to one of the main goals of the project, namely to collect new words for the Oxford English Dictionary. This way we often observe users tagging their posts with #OED as they expect an expert or researcher to come into the thread and confirm or reject the word as a new possible entry to the dictionary. As most users outside university libraries do not have access to the paid service of the OED, they are unable to confirm their finding on their own, and thus turn to the forum. As we saw in the previous section on external resources, we found numerous links to open dictionaries, indicating that some users also try out to find out the meaning of newly found words.

The #paper hashtag has an interesting dynamic of interaction. It was initiated by a super-user (who later became moderator) who noted that one author used different qualities of paper for different letters, reserving expensive and fine paper only for important messages. The same hashtag was later used by researchers, in particular one called Elaine Leong who a few days later announced “I’m doing a project on paper in recipes”. As this was written in the forum, many more users begun directing their messages towards her username, @elaineleong (see RQ 4), as soon as they found examples of paper in historical recipes in their transcriptions. So, in other words we can detect an interesting form of collective knowledge production between researchers, super-users and moderators, which is interwoven with the aid of a hashtag, and becomes a popular one used by researchers and users alike, even though it was not used as a subforum from the start of the project.

The hashtag #womanwriter(s) was first mentioned in a post by a moderator, but was then frequently used by super-users in the first months of the project. In a similar style as with the #catholic(s) hashtag, the researcher Victoria van Hyning makes frequent use of the hashtag more than a year later, in summing up the material collected. This way the data collected in the first phase becomes accessible to research as the hashtags then can be used to find back to those manuscript that are of particular interest.

### 2. Volunteer-driven hashtags

Volunteer driven hashtags include #medical, #medicine #recipe, #bleedthrough, #latin, #cooking, #food, #letter. These have in common that researchers produce less than 10% of the tags. Instead it is mainly the super-users that produce #letter(s) and #bleedthrough and with the others the main contribution stems from super-users and active users. None of these hashtags have their own subforum, but have instead grown dynamically from the heavy use of them by non-researchers. In Figure 4 we see the #medical hashtag to which researchers only make four contributions, while the bulk of the tagging is performed by super-users and active users. None of the researchers indicate that they are specifically doing research on medicine in Shakespeare’s time. Still, the hashtags #medicine and #medical continue to be widely used by non-researchers, indicating that this is an important topic of interest to them.



**Figure 4**: Timeline for the use of the #medical hashtag. Green - Researcher, Pink - Moderator, Lightblue - Super-user, Grey - Active user, White - Casual user.

As a conclusion, we see the volunteer-driven hashtags having in common that they describe phenomena unforeseen by the researchers, which instead are taken up by volunteers. There are no sub-forums for these hashtags, which suggests that the frequent texts on medicine were not expected or deemed interesting by the project creators. They are, however, of great interest to the volunteers, who invent and use these tags frequently.

## RQ 5 - Use of @-messages

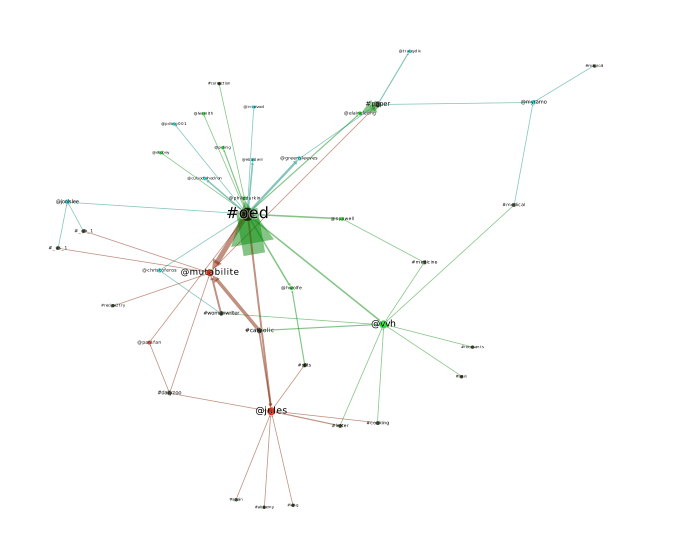
The Talk forum software used in Shakespeare’s World allows so-called “pings” (@-messages) to alert users about discussions. By writing the @-symbol in front of a user name, discussants can insure that this user will be alerted as they log in. Although this feature is used to some extent by all user groups, the most active users of pings are the researchers on the platform who were the senders of 1998 and receivers of 1413 pings.

|  |  |  |
| --- | --- | --- |
|  | Pings sent | Pings received |
| moderators | 992 | 643 |
| researchers | 1998 | 1413 |
| super-users | 471 | 666 |
| active and casual users | 171 | 910 |
| Total | 3632 | 3632 |

**Table 4:** Pings (@-messages) sent and received by user groups N=3632.

As shown in Table 5, moderators more often send pings than receive them. This was somewhat surprising, since we expected moderators to often answer questions. While this is sometimes the case, moderators are also very active users, and include in their moderator role to ping users who might know more about something, or thank users who have made a contribution. Another surprise is the distribution among researchers. Here we also find a higher frequency of sending pings than receiving them. Partly this is because at least one researcher behaves almost like a moderator, pinging users that might know more, or, as is often the case with all researchers, they frequently thank the users when they have made a contribution. Super-users receive more pings than they send, and this is also the case for active and casual users. Especially in the latter case, the large number of active and casual users, are receiving individual thank-you messages and answers to questions.

If we extract both hashtags and @-messages from each post, we are able to study the co-occurrences of hashtags and @-messages from a network perspective. Here we took each post containing at least one hashtag and one @-message and created a directed network from the hashtag to the pinged user. This way we are able to visualize a cluster, which suggests a particular distribution of topics.



**Figure 5.** Indegree network - hashtag to username. The data was extracted by matching posts in which there were both hashtags and pings. We excluded such co-occurrences if they only happened once in order to reduce data (this effectively excluded contributors below super-users). Moreover the Force Atlas 2 algorithm (Blondel et al. 2008) only concentrates nodes that are connected more than two times to the center of gravity to which the figure was limited. Legend: Green - researcher, red - moderator, magenta - superuser, black - hashtag.

In Figure 5 we can see how the #OED hashtag appears as a center of gravity, with the majority of these messages being directed towards @PhilipDurkin. Philip Durkin is a researcher working at the OED, hence this is unsurprising. However, many other moderators and researchers also receive #OED-tagged messages, indicating that the discovery of new words is not confined to a single gatekeeper, but attracts a wider community of users and discussions. Almost all researchers are pinged in connection with the #OED hashtag, proposes that this hashtag serves as an important communication point between researchers and non-researchers. However, also the two active moderators are actively being pinged in frequently, equally often as active researchers.

Another type of cluster is formed between the researcher Victoria van Hyning (@vvh), who has frequent pings with the two active moderators, especially on the hashtags #catholic and #womanwriter, but also on several more. In other words, here we find an even more genuine example of researcher-volunteer knowledge exchange. Another case of where this takes place is on the #paper hashtag, where Elaine Leong (@elaineleong) is pinged according to her research interest of paper-use in recipes. Finally, there is a case of a moderator-only hashtag, #dailyzoo, in which moderators ping each other exclusively.

In total we conclude that researchers are the most commonly pinged users, followed by moderators and super-users. In many instances, researchers are pinged because they are asked for their expertise, but such expertise has sometimes been achieved also by moderators and super-users. Among the 15 most pinged users, 7 of the total 10 researchers in the team appear. Moreover, since moderators and super-users are very active in transcribing texts and bring these to the forum, they are often thanked by the researchers who show their appreciation for their volunteer work by pinging them.

## Conclusion

One of the characteristics of online environments is that they allow to organize information through tagging, pinging and linking. Table 4 shows an overview of these features in Shakespeare’s World Talk, indicating that the different user groups make use of them in different ways.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| User roles  (RQ 1) | Number of posts | Tagging (RQ 3) | Pinging @ (RQ 4) | Linking http (RQ 2) | Indirect linking (RQ 2) |
| moderator | 3114 | 351 (11%) | 822 (26%) | 410 (13%) | 53 (2%) |
| researcher | 2132 | 254 (12%) | 1156 (54%) | 255 (12%) | 103 (5%) |
| super-user (>100 posts) | 4226 | 1542 (36%) | 371 (9%) | 186 (4%) | 78 (2%) |
| active user (>10 posts) | 1173 | 410 (35%) | 94 (8%) | 89 (8%) | 50 (4%) |
| casual user (<=10 posts) | 805 | 135 (17%) | 40 (5%) | 32 (4%) | 21 (3%) |
| total | 11450 | 2692 (24%) | 2483 (22%) | 972 (8%) | 305 (3%) |

**Table 5**: Summary of the non-linear forum organization by user groups (N=11450) and research questions. Hashtags, pings and links. The percentages are calculated in relation to the number of posts per user group.

This reveals another dimension of a citizen science project, which primarily has the purpose of transcribing texts. The different uses of the forum means that the users also attain different roles, which means that they fulfill different functions with regards to knowledge production. On the most general level we find different patterns of online behaviour among researchers, moderators, super-users and other users. It is the moderators and super-users who bring out new texts for discussion, as they are the ones doing most of the transcription work that ends up as issues on the forum. The researchers generally only respond once an issue has been posted, and then they often generously provide contextual information and means of understanding the texts. They also tag the threads and thank the volunteers for their contribution, and they often get pinged into a conversation as many users expect more information about the sometimes difficult challenge of understanding a historical manuscript.

## Discussion

Moderation is an important task to any forum. In our study we have shown that the designated moderators are investing a lot of time in tagging and pinging the discussion in order to bring the right person to the right issue. Also, some researchers may display this sort of moderator behaviour, as they know the project very well and link to resources or other researchers to expand a discussion. We distinguished three somewhat different linking behaviours among moderators, researchers and super-users. While researchers have a quite typical academic way of making references to sources outside the project, linking to research and other repositories, the use of open sources such as Wikipedia is more common among super-users and active users. Also, when there is limited access to a resource, for example the Oxford English Dictionary, users often ping a researcher with proper access.

We found that the use of hashtags was an important feature of the forum, especially for improving the search functionality and thus systematising the forum information. However, since anyone can write any hashtag without restrictions (just like on Twitter or Instagram), the issue of whether to control hashtags or not appears. If not controlled, there is a risk of overlap, for example between #medicine/#medical, #recipe/#cooking/#food, etc. However, one of the seeded tags that was introduced by researchers, #Recipes2Try, was not picked up by the users as much, perhaps because it is hard to type, or because most recipes are not recommended to try at all. Instead the volunteers used #recipe, #cooking and #food to further differentiate the findings, and perhaps making it easier to tag the content.

Summing up, we propose that hashtags and pings afford an interesting mode of organization of the knowledge produced in Shakespeare’s world. The free use of hashtags enables the emergence of new contrasts in the disseminated material, where new phenomena and discoveries can be made. The ping function, in turn, enables the formation of sub-communities of interest along certain topics, which concentrates the issues discussed and allows for special expertise to be developed and allocated at the right spot. It also creates a direct link between researchers and volunteers, which has led to some quite interesting collaborations in knowledge creation outside the regular constraints of academic humanities research.

Based on these conclusions, we suggest that VCS projects consider the following issues with regards to collaborative forums and knowledge sharing spaces. Firstly, we have shown that forums play an important role for understanding and interpreting the designated task (in our case transcription). These interaction spaces do more than just motivate people to do more citizen science, they also provide an interface between data and knowledge, where volunteer contributors can both create and immerse themselves in knowledge practices. Secondly, we have also shown the importance for researchers to respond and share their expertise with volunteers, and researchers who do that get rewarded with highly motivated citizens as co-researchers in connection to their projects. Finally, we have also seen how new phenomena are being discovered by volunteers, something which should be valued highly and credited properly.

## Acknowledgements

The authors wish to thank Dr. Victoria van Hyning for providing the forum data for the Shakespeare’s world forum and for generously sharing contextual information about how the project has developed since it was launched.

## References

Bastian, M., Heymann, S., Jacomy, M. and others, 2009. *Gephi: an open source software for exploring and manipulating networks.*

Blondel, V.D., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E., 2008. Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): P10008.

Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V. and Shirk, J., 2009. Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy, *BioScience*, 59(11): 977–984. DOI: https://doi.org/[10.1525/bio.2009.59.11.9](https://doi.org/10.1525/bio.2009.59.11.9).

Franzoni, C. and Sauermann, H., 2014. Crowd science: The organization of scientific research in open collaborative projects, *Research Policy*, 43(1): 1–20. DOI: https://doi.org/[10.1016/j.respol.2013.07.005](https://doi.org/10.1016/j.respol.2013.07.005).

Golder, S.A. and Huberman, B.A., 2006. Usage patterns of collaborative tagging systems, *Journal of Information Science*, 32(2): 198–208. DOI: https://doi.org/[10.1177/0165551506062337](https://doi.org/10.1177/0165551506062337).

Hedges, M. and Dunn, S., 2017. Google-Books-ID: SNvWDgAAQBAJ. *Academic Crowdsourcing in the Humanities: Crowds, Communities and Co-production*. Chandos Publishing.

Letierce, J., Passant, A., Breslin, J. and Decker, S., 2010. *Understanding how Twitter is used to spread scientific messages*.

Liberatore, A., Bowkett, E., MacLeod, C.J., Spurr, E. and Longnecker, N., 2018. Social Media as a Platform for a Citizen Science Community of Practice, *Citizen Science: Theory and Practice*, 3(1): . DOI: https://doi.org/[10.5334/cstp.108](https://doi.org/10.5334/cstp.108).

Luczak-Roesch, M., Tinati, R., Simperl, E., Kleek, M.V., Shadbolt, N. and Simpson, R., 2014. *Why Won’t Aliens Talk to Us? Content and Community Dynamics in Online Citizen Science*. In: 4 June 2014 United States. p. 10.

Ponciano, L. and Brasileiro, F., 2014. Finding Volunteers’ Engagement Profiles in Human Computation for Citizen Science Projects, *Human Computation*, 1(2): . DOI: https://doi.org/[10.15346/hc.v1i2.12](https://doi.org/10.15346/hc.v1i2.12).

Reed, J., Raddick, M.J., Lardner, A. and Carney, K., 2013. An Exploratory Factor Analysis of Motivations for Participating in Zooniverse, a Collection of Virtual Citizen Science Projects. In: *2013 46th Hawaii International Conference on System Sciences*. January 2013 pp. 610–619. DOI: https://doi.org/[10.1109/HICSS.2013.85](https://doi.org/10.1109/HICSS.2013.85).

Sinclair, J. and Cardew-Hall, M., 2008. The folksonomy tag cloud: when is it useful?, *Journal of Information Science*, 34(1): 15–29. DOI: https://doi.org/[10.1177/0165551506078083](https://doi.org/10.1177/0165551506078083).

Tinati, R., Van Kleek, M., Simperl, E., Luczak-Rösch, M., Simpson, R. and Shadbolt, N., 2015. Designing for Citizen Data Analysis: A Cross-Sectional Case Study of a Multi-Domain Citizen Science Platform. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI ’15. 2015 New York, NY, USA: ACM. pp. 4069–4078. DOI: https://doi.org/[10.1145/2702123.2702420](https://doi.org/10.1145/2702123.2702420).

Trant, J., 2009. Studying Social Tagging and Folksonomy: A Review and Framework, *Journal of Digital Information*, 10(1): .

Trant, J., Bearman, D. and Chun, S., 2007. The eye of the beholder: steve. museum and social tagging of museum collections. In: *Proceedings of International Cultural Heritage Informatics Meeting (ICHIM07)*. 2007 p.

Veletsianos, G., 2017. Three Cases of Hashtags Used as Learning and Professional Development Environments, *TechTrends*, 61(3): 284–292. DOI: https://doi.org/[10.1007/s11528-016-0143-3](https://doi.org/10.1007/s11528-016-0143-3).

Wiggins, A. and Crowston, K., 2011. From conservation to crowdsourcing: A typology of citizen science. In: *System Sciences (HICSS), 2011 44th Hawaii international conference on*. 2011 IEEE. pp. 1–10.

Zhang, A., Zheng, M. and Pang, B., 2018. Structural diversity effect on hashtag adoption in Twitter, *Physica A: Statistical Mechanics and its Applications*, 493: 267–275. DOI: https://doi.org/[10.1016/j.physa.2017.09.075](https://doi.org/10.1016/j.physa.2017.09.075).

1. The exact numbers of @-messages, hashtags and URLs vary between Regular Expressions engines. In this article we use the engine built into Microsoft Excel 2016, which we verified against Python3’s regexp engine. The results only differed in the range of 5-10 hits. [↑](#footnote-ref-2)
2. This hashtag is also expressed as #womenwriter(s) by 5 users with the same meaning. On one occurrence this form predates #womanwriter, in a post by Victoria van Hyning, who could be called the initiator of the hashtag as a whole. [↑](#footnote-ref-3)