Folksonomies in Crowdsourcing Platforms: Three Tensions Associated with the Development of Shared Language in Distributed Groups

Julia Bullard[1], Kevin Crowston[2], Corey Jackson[3], Alexander O. Smith[2] and Carsten Østerlund[2]

[1] School of Information, University of British Columbia, Vancouver, BC, Canada
[2] School of Information Studies, Syracuse University, Syracuse, NY, USA
[3] The Information School, University of Wisconsin–Madison, Madison, WI, USA
Contact Author: crowston@syr.edu

Abstract. Members of cooperative groups can work together more effectively if they develop a shared language, but distributed groups face barriers to doing so. To better understand how shared language can emerge in and support the work of distributed groups, we review the literature on folksonomies (a kind of shared language) in crowdsourcing systems (one type

of distributed work). The review highlights three tensions associated with the development of folksonomies in crowdsourcing. The first problem is who has the power to decide on adopted terminology. Second, different users of the language may have different needs. There might be tension if people tagging objects are not the same as those using these tags to search for content. Finally, projects need to decide when to intervene to maintain a balance between a stable ontology and the ability of the project to accommodate ongoing changes. We illustrate these considerations by comparing how they are handled in the story-sharing site *Archive of Our Own*, the citizen-science project *GravitySpy*, and the photo-sharing site *Flickr*.

Introduction

Members of distributed groups face challenges in achieving coordinated action. In this setting, the development of a shared language (i.e., specialized terminology used to describe the shared work and work situations) plays an essential role in supporting the coordination of group members' practices (Crowston and Kammerer, 1998). Without a common interpretive schema, individuals with different backgrounds may interpret situations and tasks differently, hampering collaboration and communication (Dougherty, 1992). For instance, the citizen-science project Snapshot Serengeti, in which volunteers identify the species of animals in pictures from African camera traps, would not work if every volunteer had their own names and definition of animal species. In this case, the shared language to be used is set by external authorities, i.e., the definition of species is given by zoologists to the citizen science volunteers. In many other cases, though, a shared language needs to emerge from the community's practices, e.g., the description of a novel species from the point of view of working zoologists. However, in highly distributed environments, individuals may have few opportunities for informal discussion and mutual observation that can foster the development of shared language and interpretive schemas.

To explore the bottom-up development and use of shared language in highly distributed work arrangements, we examine a specific kind of shared language, namely folksonomies, in a specific kind of distributed work, namely crowdsourcing. Both terms, crowdsourcing and folksonomy, are portmanteaus, bringing together crowd and sourcing and folk and taxonomy respectively. The notion of crowdsourcing was popularized by Howe (2010), who defined it as outsourcing of functions to an undefined (and generally extensive) network through an open call. Folksonomy refers to user-generated classification methods in which users collaboratively generate open-ended labels (also referred to as "tags") to categorize content (e.g., web resources, photos, citizen science data, and other online material to facilitate retrieval and use (Bullard, 2019; Noruzi, 2006; Trant, 2009). The decentralized nature of crowdsourcing and the bottom-up nature of folksonomy development provides a lens into the development of shared language in distributed work.

The specific goal of this paper is to review different strategies (i.e., what decisions to make and when) for supporting the development of folksonomies to

better understand the impact of different decisions for promoting the development of a shared language in crowdsourcing settings. The paper's novelty is to take a project-management perspective by identifying three key questions about how a project is run that need to be addressed and discussing factors that weigh on those decisions. We start by reviewing the existing literature on crowdsourcing and folksonomies as a basis for identifying specific strategic questions to be investigated.

Literature review

Crowdsourcing and shared language

Crowdsourcing is a model for collaboration that uses information and communication technologies to access human effort to solve complex problems. In different settings, crowdsourcing has been called peer production, user-powered systems, user-generated content, collaborative systems, community systems, social systems, social search, social media, collective intelligence, wikinomics, crowd wisdom, smart mobs, mass collaboration and human computation (Doan et al., 2011). Crowdsourcing has been applied in a wide variety of business and academic domains. For instance, in academic research, working scientists employ crowdsourcing (in this setting called citizen science) to harness the efforts of amateur volunteers towards scientific tasks including developing research questions, collecting data, analyzing data and writing up results (Bonney et al., 2009). For example, the Galaxy Zoo citizen science project asks volunteers to review images of galaxies and report on their morphological and evolutionary features (Fortson et al., 2012).

In many crowdsourcing projects, participants work independently. However, when the tasks carried out by different workers are interdependent, a common language is essential to support effective coordination. The literature on shared language reveals that "language practices are instrumental in creating the norms of behavior of particular online groups" and "these norms function to provide sociability, support, information and a sense of collective identity" (Lam, 2008). When shared language is achieved, it provides benefits such as improved language quality, conversational grounding, retention, community sociability and support, and a sense of collective identity (Lam, 2008).

While a shared language is a common need, the characteristics that define many crowdsourcing projects make it challenging to achieve. First, the crowd's composition may influence a project's ability to achieve a shared language. As most crowdsourced projects recruit members through open calls, the resulting "crowd" may have diverse educational attainment and expertise. For instance, a Galaxy Zoo volunteer without the requisite training or educational background may struggle to comprehend the language used by working scientists or more advanced contributors, limiting the types of tasks that they can conduct.

Second, there can be issues with the evolution of a project's language. Participants may learn the language at a different pace, meaning that the language is discordant across the project. Newcomers who join a project during its later stages may

struggle to make sense of the project's emergent language. As Feinberg (2006) argues, aggregates of individual language and democratic decision-making processes are not necessarily 'cohesive' collaborations. Feinberg points out that a project's language can become increasingly complex and incoherent as data increases and individual motives and collective goals are accounted for. These challenges will be exacerbated if there are no training materials from which to learn the language.

Finally, the governance structure adopted by projects may lead to uncertainty about the appropriate source of the language to be employed in the community. While some projects may have a hierarchical governance structure in which a few pre-selected persons fill leadership roles, many members can evolve their responsibilities and take on new roles to sustain the community. If decision-making power about the appropriate language is decentralized, contributors may not know from whom to take linguistic cues or even perceive a need to adapt their language. As a result, projects may lack processes for adjudicating new language in the project.

Folksonomies

A folksonomy is a user-generated classification and information-retrieval method with three characteristics (Peters and Weller, 2008; Bullard, 2019; Trant, 2009). First, folksonomies constitute collaboratively-generated work by large and distributed groups as part of their ongoing tagging of objects, collections and searches in Internet-based platforms. There is a grass-roots element to the production and use of folksonomies where the tagging practices and resulting folksonomies service emerging communal purposes. Second, as a common use of tags is to facilitate search and retrieval, tags need to serve the needs of both individuals trying to organize their own materials and searchers looking for materials, so content created or tagged by one user can be found by others. Third, the set of tags is often open-ended, so the size to which and speed with which folksonomies grow can defy the temporal rhythms of traditional knowledge organization systems.

Folksonomies can be contrasted to formal knowledge organization systems (KOS) and computational approaches to information organization (Bullard, 2019). On the one hand, the controlled vocabulary associated with a formal KOS is developed, applied and controlled by experts and is typically structured and stable (Hjørland, 2008). On the other hand, computational approaches generate search terms algorithmically, which reduces the need for expert labor. These approaches benefit from large corpora for analysis (Zhitomirsky-Geffet et al., 2016). However, a social process is still needed to transfer the generated terms into practice.

Enabling online users to contribute labels or tags for personal and shared information organization and retrieval has gained attention in a range of fields. The library, museum and archives community have experimented with folksonomies in an effort to broaden the voice and engagement of their community and stay current (Trant, 2009; Yi and Chan, 2009; Stewart and Kendrick, 2019; Lu et al., 2010). These communities also have studied existing folksonomies to produce accurate

metadata to apply to collections (Dieckman, 2022). Education has also shown interest as part of an effort to improve the searching of networked resources in ways that support both collective and personalized uses (Miller, 2005). Folksonomies have also been studied for geo-spatial mapping (Mocnik et al., 2017), healthcare (Smith and Wicks, 2008), multimedia (Zheng et al., 2016), ecommerce recommendation systems (Mao et al., 2021) and finance (O'Leary, 2015). This broad interest has led to a proliferation of overlapping terms for the concept. Trant (2009) finds other terms describing aspects of this phenomenon, including social tagging and social classification (Landbeck, 2007), community cataloging and cataloging by the crowd (Chun and Jenkins, 2005) and ethnoclassification (Walker, 2005).

Folksonomies in crowdsourcing

Next, we discuss how folksonomies work in crowdsourcing settings to illuminate the processes by which shared language can emerge in distributed work settings. Folksonomies emerge from individual and communal practices that involve the tagging of individual objects and the creation of object collections. Figure 1 summarizes the core elements in this process as conceptualized by Jackson et al. (2018). Starting in the upper left, by tagging (or tagging) individual items, participants seek to assign them meaning. For instance, an author may seek to tag the genre or topic of a story they have read. As they tag multiple items, they may seek to be consistent in the tagging so that the tags can serve to connect related objects, e.g., all of the cozy mysteries in a particular setting. Through this process, individuals gradually create their own 'personomies', that is, a categorization system unique to their own practices (upper right). Multiple purposes can drive the development of personomies, as individual tags do not necessarily refer to the content of the objects but can also denote author, origin, data form, work process or context-dependent characteristics of the object salient to the individual (Macgregor and McCulloch, 2006). Ames and Naaman (2007) find that individuals have four main motivations to tag items in social tagging systems, including self organization, self communication (i.e., to reflect and keep track of daily life), social organization (i.e., creating a shared collection) and social communication (i.e., sharing opinions about a specific item). For instance, it is not uncommon to find self-directed tags, such as "to read."

	Single Objects	Collections of Objects
Individual	Object tagging $\rightarrow$	Personomy
	$\downarrow$	$\downarrow$
Collective	Object-tagging consensus $\leftrightarrow$	$\leftrightarrow$ Folksonomy

Figure 1. Folksonomy creation process.

At the collective level (lower left quadrant in Figure 1), multiple users examining an object may strive to reach an agreement about the most appropriate tags to describe it. Agreement on tagging may come through discussions about a particular object or

more directly when one individual mirrors the observed practices of others and so applies the same tag. When visible to others, individual tagging practices may serve as a guide for other's practices.

Finally, folksonomies (the lower right quadrant) take shape as personomies and collectively-agreed-upon tags accumulate over time. Discussions about these tags may further facilitate consensus by allowing individuals and groups to base their tagging on previously-shared tags. Contrariwise, a developed folksonomy provides the basis for group members to coordinate their tagging and improve retrieval of relevant objects in their searches, meaning that the flow between the lower left and right quadrant goes in both directions.

Unfortunately, the free-tagging practices feeding the production of folksonomies rarely achieve a coherent and structured ontology. Instead, most folksonomies are Peters and Weller (2008) called "tag gardens," each tag like a plant growing wild. A few tags may receive much attention, but others still proliferate (i.e., a long-tail distribution (Munk and Mørk, 2007)), yielding an unruly and overgrown garden. Research has revealed specific shortcomings of folksonomies (Al-Khalifa and Davis, 2007; Lee and Schleyer, 2012; Aurnhammer et al., 2006; Guy and Tonkin, 2006; Spiteri, 2007). Terms may be spelling variants, abbreviations, initialisms or acronyms, synonyms, compound tags where two words are joined or duplicate generic and specific names for a concept. Terms may also be ambiguous by using neologisms, slang, jargon (Spiteri, 2007) or even tangential humor (Price, 2019). A person's or group's use of tagging strategies may change over time (Begelman et al., 2006) or subcultures may form as smaller groups develop vocabularies to serve their particular interests (Rafferty and Hidderley, 2007).

A number of techniques have emerged to manage the messy nature of emerging folksonomies and overcome some of their shortcomings. Peters and Weller (2008) refer to these as "gardening techniques" that re-engineer folksonomies to make them more productive for broader use. These interventions target different parts of the process (see Table I). First, some strategies attempt to guide individuals in tagging (upper left side of Table I). For instance, as a user enters a term, the system can spell check or auto-complete suggestions from the existing folksonomy (Bullard, 2019; Munk and Mørk, 2007). Price (2019) describes a *post hoc* disciplinary technique generated by human domain experts judge tags based on literary, scientific, user or ethical "warrant." Some sites use word clouds to call attention to popular tags to help guide individuals in their selection process, though this approach risks promoting tagging driven by imitation as opposed to thoughtfulness (Munk and Mørk, 2007). Other strategies include computationally-derived suggestions for tags (Razikin et al., 2011), e.g., using centrality in a network of tag co-occurrences (Price, 2019).

Second, strategies can attempt to shape the personomies that feed many folksonomies (upper right quadrant). Curiously, we find few examples of such functionality, beyond simply listing tags used and enabling searches for items that use a tag. One could imagine allowing individuals to easily review and curate their own collection of tags, e.g., to easily change the name of a particular tag. Likewise, it might be useful for people to compare their personomies with other personomies

	Single objects	Collection of objects
Individual	• Autofill • Spell check • Knowledge Organization System (KOS) suggestions • Computationally derived suggestions	• Personomy management

Communal	• Object tagging consensus reached in Talk and Discussion forums embedded in the tagging process	• Expert curated
		• Computational processes:
		  – Word Clouds and other visualizations
		  – Computational techniques to find clusters, tag-networks, readable dictionaries, etc.
		• Community curated by:
		  – Any contributor
		  – Curators identified by algorithmic reputation
		  – Volunteer curators selected and trained

Table I. Gardening techniques and the focus of their interventions.

or a shared folksonomy.

Third, a number of efforts strive to facilitate communal consensus as part of the classification process for individual contributions (lower left side quadrant). For instance, on the citizen science platform Zooniverse, participants can see what tags other users have applied to particular objects and engage in a conversation about their tagging in discussions of tags in the Talk and Discussion fora.

Finally, some gardening techniques intervene at the folksonomy level and approach the whole folksonomy as a raw material from which to derive a structured knowledge-organization system (lower right-side quadrant). The folksonomy review can be driven by experts, computational techniques or the user community. An expert-driven approach employs professional knowledge organization designers to review the folksonomy and revise descriptors (Syn and Spring, 2013). Computational techniques attempt to transform folksonomies into coherent taxonomies and ontologies. Clustering, matching algorithms, machine-readable dictionaries and tag networks have all been explored as possible ways to process folksonomies (Tsui et al., 2010). Others take existing tags and apply machine learning algorithms to find similarities *post hoc* (Mao et al., 2021; Dieckman, 2022; Zhou et al., 2023). These computational approaches worked well with large corpora and ongoing changes to folksonomies and reduced the need for expert labor (Zhitomirsky-Geffet et al., 2016).

Finally, on some platforms, the user community structures unruly folksonomies. Bullard (2019) describes three empirical cases that, in different ways, divide up the work to make changes to the folksonomy.

- In the project *LibraryThing*, users not only develop a database of books but also curate a folksonomy of tags applied to the books. Any user can edit tags, for instance, by combining synonymous tags, separating wrongly combined tags or calling a vote to reach a communal decision on important tags. *LibraryThing* keeps track of tag names' provenance so users can review the history of a term and what changes have taken place in the past. Equally important, the revision process does not change the original tags on a book or user's page. Only on site-wide retrieval do the combinations of terms come into play. In other words, personomies and folksonomies maintain separate but interrelated lives.
- In *Stack Overflow*, a question and answer site for programmers (with spinoffs for many other domains), incoming questions can be given tags to describe the topic of the question to help steer them to the right person to answer them. While any phrase can be used as a tag, terms can be noted as synonyms so that divergent terms get replaced. The system assigns the right to edit the folksonomy of tags to users with high algorithmic "reputation" whom the system has judged as good contributors. Members of this group can decide to merge or separate tags (Bullard, 2019, p. 647).
- A more selective approach to who gets involved in folksonomy editing can be found on *Archive of our Own* (AO3), a site that offers a noncommercial and nonprofit host for fanworks (e.g., stories set in fictional universes using those characters, e.g., a story set in the world of Harry Potter). On this site, a few hundred volunteers who have completed a recruitment and training process curate the site's folksonomy. Once appointed, these "tag wranglers" have significant autonomy to manage their assigned section of the folksonomy (Bullard, 2019, p. 647).

Three tensions in folksonomy development

From the literature review, we identify three tensions that arise when promoting development of a folksonomy in a crowdsourcing setting that go to the core of shared language development. A first tension regards who has the power to decide on adopted terminology. As noted above, in many settings, power is centralized and an authoritative shared language can be imposed on group members. In the absence of authority, different users may strive to control the language used. Or, it may be that there is no need for consensus on tagging meaning that there is no movement from individual to collective in Figure 1 beyond simple aggregation.

Second, in distributed work settings, different workers may have different needs for the shared language. In the setting of folksonomies for crowd work specifically, we see a tension between the needs of users who tag content to classify it for their own purposes and those who use the tags to search for content, i.e., between individual and collective needs (upper vs. lower portions of Figure 1).

A final related tension has to do with temporality, i.e., when in the work process shared knowledge plays a role. Does it make more sense to seek to apply shared terminology at the time of tagging an object (left side of Figure 1) or later in the process when there is a collection of tagged objects (right side of Figure 1)? These decisions relate to the needs for stability and dynamism. On the one hand, an advantage of folksonomies is that they can be dynamic, changing as work practices shift and thus accommodating a changing situation. On the other hand, changes can be disruptive to the use of folksonomy terms, so users may prefer some degree of stability.

These tensions led to the three strategic decisions we address: (1) How does a project allocate authority to make decisions about the folksonomy and support that authority? (2) How does a project weigh the support for those who develop terms (i.e., tagging objects) versus those who use the tag? (3) When does a project intervene in the work process to promote the development of a shared language?

Method

We addressed our research questions through a multiple comparative case study of three online communities: Gravity Spy[1], Archive of Our Own (AO3)[2] and Flickr[3]. These three were chosen because they span a range of practices related to folksonomies, as well as addressing different users and different scales.

- Gravity Spy is a citizen-science project hosted on the Zooniverse platform. In this project, volunteers classify "glitches" or noise events from the Laser Interferometer Gravitational-Wave Observatory (LIGO). Most volunteers classify glitches into classes identified by LIGO scientists but advanced volunteers seek to identify novel classes of glitches, tagging them with new terms.
- AO3 is a fanwork sharing site, introduced above. Contributors of stories tag them to identify the content while tag wranglers connect the applied tags to a site-wide folksonomy to facilitate searching for stories.
- Flickr is a commercial site hosting photographs and some video. Photos can be grouped into albums but also tagged by the poster, e.g., for content, setting or some other purpose and these tags can be searched for.

For each site, we provide contextual illuminations of participants' practices through virtual ethnography (Hine, 2008; Østerlund et al., 2020). Virtual ethnography adapts traditional ethnographic methods, such as participant observation and in-person interviews, to studying online communities. Each author has been a participant in one of the projects for several years and has contributed to the project classifying and searching for data, stories or pictures, and reading and posting in discussion fora. As participants, we created user accounts, completed requisite training (if any), created tags, and read and contributed to project discussions over the course of the project. For instance, one author has served for years as a tag wrangler on AO3. We

---

[1] https://www.gravityspy.org/
[2] https://archiveofourown.org/
[3] https://flickr.com/

used our position as observers to build knowledge about how volunteers and site users engage with folksonomies. We analyzed the process we identified to determine how they mapped to the categories in the theoretical framework. In other studies (e.g., Bullard, 2019; Jackson et al., 2018) we interviewed volunteers about their work on folksonomies. Those interviews provide additional background for our investigations rather than being a primary source for this article.

Findings: Tensions and decision points in the development of shared language

From our participant observation, we have identified user practices and system features that indicate how the three different project have addressed our three questions. We discuss each in turn.

Allocation and support of power

Gardening a folksonomy can be trusted to different participants in the community. Assigning gardening tasks allocates the power to shape the communal language. For instance, some projects take a laissez-faire approach where anybody can edit the language, others choose a democratic strategy involving voting and others have a small group of experts who make the final decisions on shared terminology. However, the allocation of power needs to be supported by technical features for exercising it.

In *Gravity Spy* there are two sets of tags and power is exercised differently over the them. Most volunteers classify glitches into the classes defined by the science team, as shown in Figure 2. The science team retain the final decision on what terms get included in the recognized list of tags. If volunteers think a new glitch class should be added, they need to go through an approval process that includes the submission of a document detailing the glitch morphology, tags associated with this glitch on Talk and example images. Science team members then review the proposal to investigate the glitch class further before they decide if it should be included in the classification system (Zevin et al., 2024).

Separately, advanced volunteers tag glitches that do not seem to belong to an existing classes, creating new tags in Talk as they do so. These efforts take place on the Talk board, as shown in Figure 3, and are largely self-managed. A few moderators, selected from the active volunteers by the science team, monitor the discussion and help answer questions and moderate conflicts (what Crowston and Fagnot (2018) would describe as meta-contribution, contributions that enable others to contribute more effectively). However, moderators hold little power to sanction or shape other participants' tagging behaviors. Volunteers are free to develop and maintain their own personomies and to use or ignore the developing folksonomy. Nor have the project scientists organized any effort to garden these tags. As a result, the emerging folksonomy is unruly and suffers many of the shortcomings associated with open folksonomies. Furthermore, the Zooniverse platform does not provide functionality to
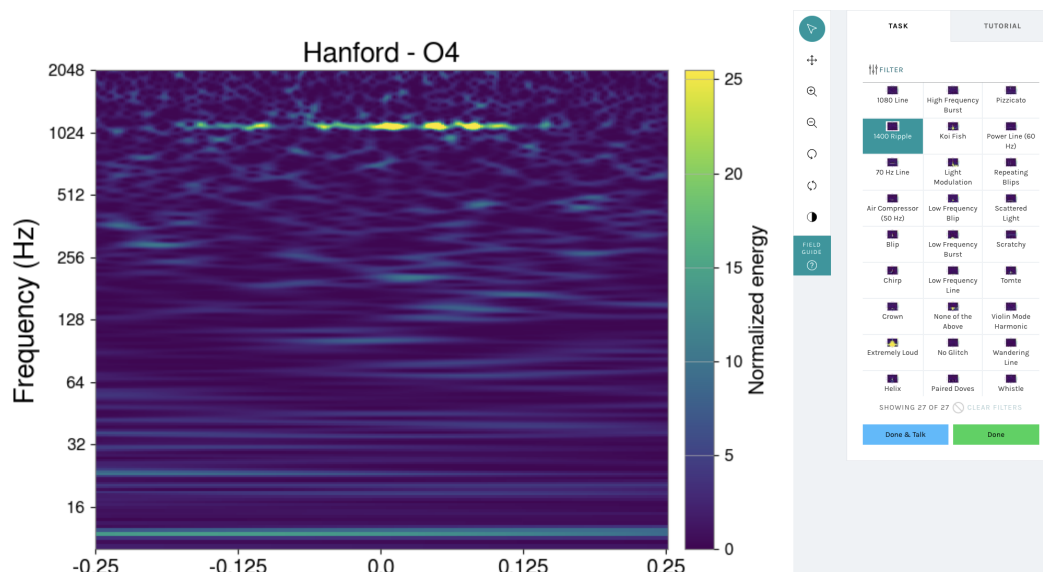
Figure 2. The Gravity Spy classification interface, with a glitch spectogram on the left and the scientist-approved tags on the right.
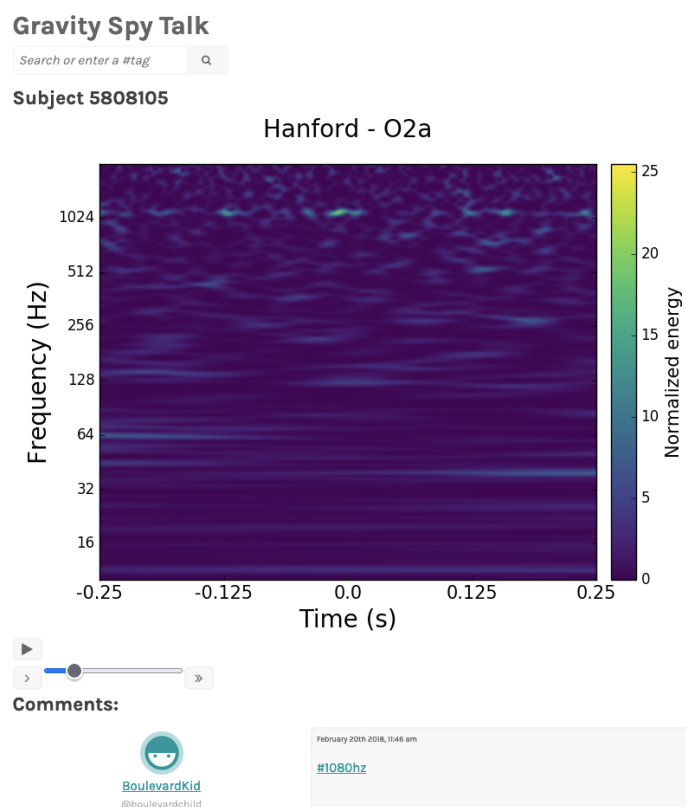

Figure 3. The Talk page for a particular subject showing how a tag (in this case, #1080Hz at the bottom right) can be applied to label glitches that do not seem to fit an existing class. The Zooniverse platform supports searches for discussions including a tag, e.g., by clicking on the link in the comments.

manage a folksonomy. For instance, even if moderators were to settle on a preferred tag, there is no easy way to update tags previously applied to objects, to indicate that tags are synonyms or to mark a tag as preferred or not.

AO3 has bifurcated the power to tag individual objects and to manage the folksonomy. Individual users generate and own the content and the tags they apply and any personomies they develop are not edited by others. However, at the folksonomy level, approximately 200 "tag wranglers" (out of 6,698,000 users) have power to merge, separate and hierarchically organize tags to develop a coherent folksonomy that supports searching. The system provides functionality to support this process, e.g., to indicate which terms are synonyms and which should or should not be auto-completed when tagging, as shown in Figure 4. Further, the project has developed extensive policies to guide the process.

In *Flickr*, contributors are the owner of their data and generally are the only ones who have authority to tag their photos, as illustrated in Figure 5. They may choose popular tags to make their photos more readily searchable by the community or instead use tags with personal meaning. There is no authority nor any specific process or system support to curate the folksonomy or to clean up the tags to maintain controlled vocabularies. In the framework of Crowston and Fagnot (2018), there is no possibility for meta-contribution to Flickr.

Supporting taggers or searchers by developing consensus

Supporting the development and use of a folksonomy must consider who tags objects versus who uses tags to search for objects. If it is the same person tagging and retrieving, different gardening techniques may be more appropriate than if tagging and retrieving are done by different people.

In *Gravity Spy*, most users classify glitches into classes defined by the science team, supporting the science team's use of the tags for searching for instances known types of glitches. The Zooniverse platform makes building such a classification system straightforward. In contrast, the tags created by advanced volunteers for novel glitches mostly serve these volunteers, as science team members do not engage with them unless called upon to do so. There is some communal usage, e.g., one advanced volunteer looking for images tagged by another, but such activity is limited. The focus for building consensus is the development of proposals for new glitch classes, which a handful of active volunteers undertake through discussion on the "New Glitch Classes" Talk board, which has had 18 contributors out of 33,739 volunteers to the project.

Archive of Our Own (AO3) primarily faces outward. Authors tag contributions and volunteers curate the folksonomy to increase the retrievability of contributed content. For readers, tags are both a retrieval and filtering tool; given diverse tastes in the community, an important application of tags in the folksonomy is the application of the Boolean NOT to eliminate unwanted topics. Similarly, avoiding the gardening technique of editing the tags that individual users assign to a contribution is consistent with the community value of creative freedom. Instead, the tag wranglers attempt
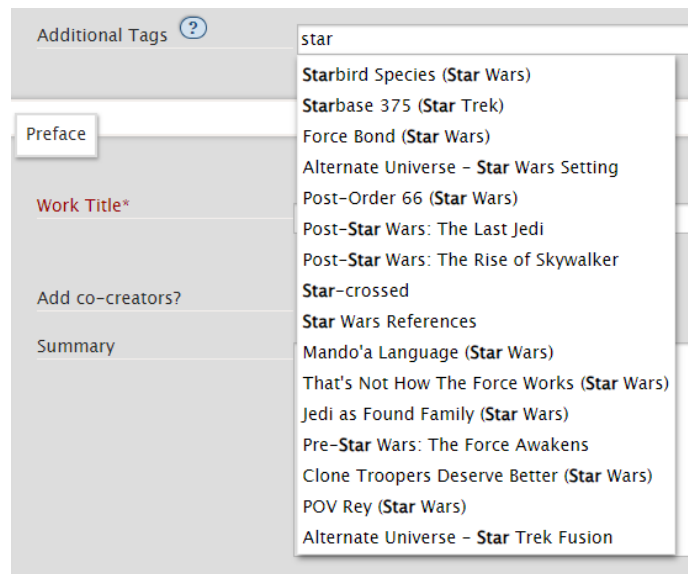
Figure 4. The AO3 new work interface showing autofills populated by tag wrangler-approved tags.
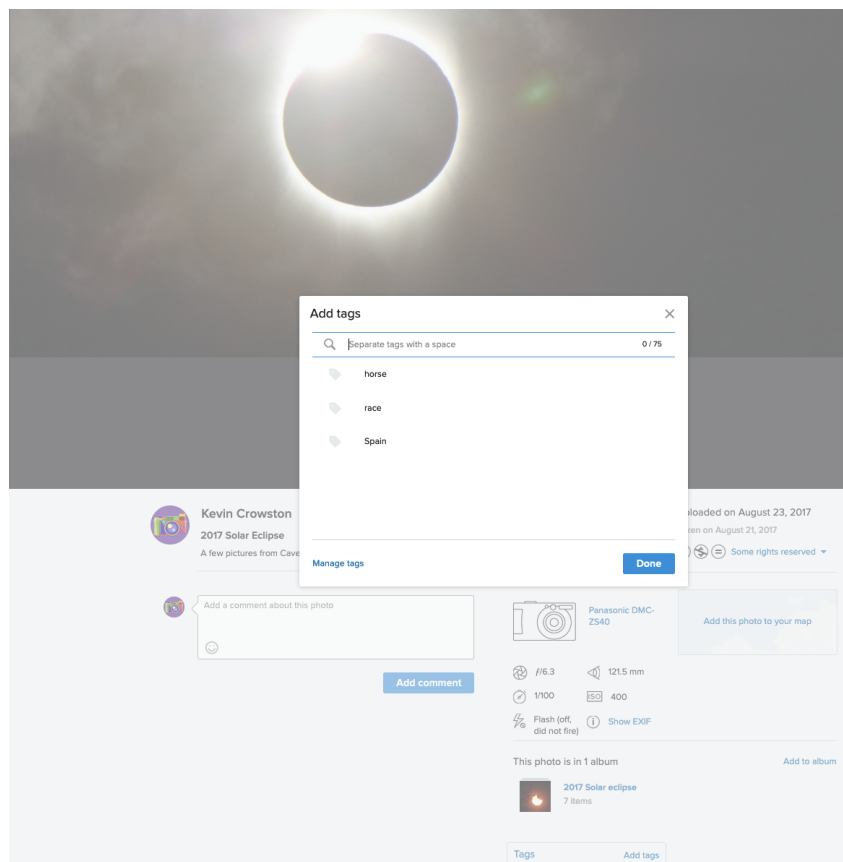

Figure 5. The Flickr interface showing how tags can be applied to a photo. The popup includes previously used tags.

to fit the contribution into the existing folksonomy by equating terms to improve retrieval. Tag wranglers make their decisions independently, but strive to identify the consensus of contributors about the meaning of tags. They can also discuss problematic issues with other tag wranglers in private discussion fora.

In contrast to the previous examples, *Flickr*, a photo sharing site, seems to primarily support individuals' tagging, though with unclear benefit. Individuals can apply tags to organize their collection of uploaded pictures, but they can also sort them into albums. There is little or no attempt to have contributors reach a consensus on tags for a specific photo or to agree on an overall organization scheme. As in Gravity Spy, one can find discussions regarding appropriate tags, but it is likely that few users have seen them or feel motivated to follow them. As a result, the support tags provide for searching is compromised, since there is promise of neither precision (finding only relevant images, since they may be tagged indiscriminately) nor recall (finding all relevant images, since they may not be tagged consistently).

When in the process and how is stability maintained?

Finally, projects face a choice of when to intervene in the process to maintain the stability for the folksonomy. Specifically, a project can promote gardening techniques as new contributions are tagged or later in the process at the collections level.

In *Gravity Spy* project, one finds two levels of stability and temporality. In the Talk and Discussion forums, the speed of change is high and there is little or no attempt to keep a stable folksonomy. However, at the project level, changes to terminology come slowly and effort is put into maintaining a stable and lasting classification system that has been vetted not only by volunteers but also by expert science team members. Unfortunately, the review process is time consuming and so LIGO scientists have been able to examine only a few of the proposals generated. As a result, there is a struggle to keep the recognized list of glitches in line with changes in the detector that fix some glitches but yield some new ones. This being said, the team behind *Gravity Spy* have discussed if they should promote more structured tagging practices at the point where individuals tag an object in Talk. They have considered autofill, spell-check and computationally-derived suggestions that would help volunteers compare their own tags with existing tags. However, the Zooniverse system would need to be updated to support these functions.

During tagging, users in *AO3* do not seem to attempt to maintain stability. Users may infer tagging conventions from autofill suggestions, as only tags established by tag wranglers appear in such suggestions, though there is no obligation to follow them. However, stability is maintained at the folksonomy level, as a select group of trained participants links synonymous terms, differentiates homographs and establishes relationships among terms. As the system preserves users' original tags, the provenance of the terms is maintained.

In *Flickr*, when individuals tag images the system suggests previously used tags, which individuals can choose to accept or decline. In other words, there is limited effort to maintain stability for a particular user, but not across multiple users.

	Gravity Spy	Archives of Our Own (AO3)	Flickr
Power and support	• Contributors generate tag; no enforcement to use particular tags • Scientists have final say on additions to recognized terminology	• At the personomy level, contributors generate and own data and maintain their own tags • At the folksonomy level, tag wranglers are recruited and trained and given permission and support to curate the folksonomy	• Contributors choose and maintain tags; no requirement to use particular terms • There are discussions of appropriate tags, but completely informal
Tagging vs. searching (Individual vs collective)	• Tagging with known tags serves science team • Tagging in Talk mostly serves the person doing the tagging; some use by fellow volunteers	• Curation process does not change user-chosen tags • Terms are linked for retrieval by connecting tags at folksonomy level	• Can be used by both contributors and searchers but doesn't seem to serve either well • Suggests previously used tags for individual object tagging
When in the process	• Intervenes early for known tags and late for novel tags • Only a few new terms get added to the official classification system • Unstable in the short term but seeks stability in the long term	• Intervenes late in the process by managing the folksonomy • Only equates terms at the folksonomy level for retrieval • Maintain provenance of tag history	• Minor intervention early at the time of tagging objects by offering tag suggestions • Fast-changing, with no need for stability

Table II. Central tensions in the development of a shared language.

Our findings are summarized in Table II.

Discussion

The development and maintenance of folksonomies offer a window into the process of building and maintaining a shared language in highly distributed groups and in particular crowdsourcing settings. While many efforts to develop shared terminology take a top-down approach where experts promote official knowledge organization systems, folksonomies allow us to consider bottom-up approaches where the language develops over time and may serve diverse purposes. Furthermore, folksonomies allow us to determine how different crowdsourcing projects may develop different structures of legitimation and authority over the emerging ontology. Different projects face different users and objectives and so make different decisions about how to address challenges in supporting folksonomies. For each of the questions addressed above, we summarize across the projects and identify factors that influence choices.

First, who gets to decide on adopted terminology stands as an important problem. However, it is not a simple choice between a top-down or bottom-up approach. Projects like AO3 allocate power to both individual contributors and a selective group of trained participants but at different points in the process. By doing so, they give individuals free range to name their objects but allow the community to develop authoritative structures facilitating search and retrieval. Other projects like *LibraryThing* permit all participants to edit and vote on tags for the emerging folksonomy. At the same time, they carefully keep track of its provenance so decisions can be debated and reversed at a later stage if the situation or the community call for such interventions. These are just a few options for the division of authority.

There are three important considerations for questions about who manages a folksonomy. First is matching the delegated power to the technical ability to make and enforce decisions. These decisions need to go hand-in-hand, as systems evolve to support desired functions or expectations of folksonomies managers are tempered by limits on their capabilities, as in Gravity Spy. Second, supporting a more developed management process is costly, requiring a lot of mostly unseen labour to make it happen. For instance, AO3 requires continued contributions from its tag wranglers; Gravity Spy needs support and proposal review by LIGO scientists. The motivations for these contributions need to be considered. In the case of AO3, tag wranglers are volunteers, what Crowston and Fagnot (2018) called meta-contributors. They suggested that such contributors are motivated by ideology, social factors such as group identification and and the intrinsic enjoyment of the task, and that projects should recognize and reward those who take on these roles. In contrast, the LIGO scientists involved in Gravity Spy are hoping for data to improve the detector but the connection of the project to their work is at best indirect, reducing the urgency of contributing. Finally, the complexity of the folksonomy may limit who can contribute. The folksonomies in Gravity Spy and Flickr are just lists of terms but AO3's has more structure, requiring more expertise to manage.

Second, the goals of different users of the folksonomy need to be weighed. In particular, tensions can arise if the people tagging objects are not the same as those using these tags to search for content. In this situation, interests might not be aligned between the individuals tagging and the collective need to facilitate retrieval of shared objects. Some projects appear to solve this problem by maintaining two systems. For instance, AO3 allows individuals to develop and maintain their own tags. At the same time, they give a small group the authority to garden the shared folksonomy in ways that facilitate search and retrieval.

Different motivational factor may lead projects to promote gardening techniques that intervene at the individual or the communal level (top or bottom of Figure 1 and Table I). If the purpose of the tagging is driven mainly by individual needs, the project may want to intervene at the personomy level to help individuals achieve consistency in their tagging. Contrariwise, if the goal is to support the work and communication of the community, the project needs to focus on promoting consensus in the tagging situation and developing coherent folksonomies. Of course, these are not exclusive: projects may need to support both individual and collective work.

A key question here is the process for achieving consensus, if any, and what kind of system support is provided. Questions about how to achieve consensus are closely tied to questions of who is involved, since it defines what group needs to come to agreement. In the three cases, we see a range of approaches from discussion to individual decisions to no process at all. In the absence of consensus, the emergent folksonomy will be nothing more than the concatenation of individual personomies. For instance, Flickr is participatory, in that users contribute the tags, but not consensual. It is worth noting that few gardening efforts appear to target personomies and the folding of curated personomies into larger folksonomies. One can imagine projects that help participants develop coherent personomies with high value for the individual that could then serve as the basis for the curation of a folksonomy.

Finally, projects need to decide when in the process of language development they want to intervene in order to maintain a balance between a stable ontology and the ability of the project to accommodate ongoing changes. Considering the different gardening techniques, projects can either intervene at the time of tagging or wait for a folksonomy to develop before cleaning up the classification of contributions (assuming there are tools available to do so). A particular factor to consider is the needs of newcomers to the project. As a project matures, so does the size and complexity of its folksonomy. If thousands of volunteers have added their own tags with little coordination, newcomers will be overwhelmed by an unruly folksonomy with many overlapping and hard-to-interpret terms. To deal with this complexity, they need guidance on how to tag objects as they get started. Only with experience might they later develop an appreciation of the freedom to choose their own tags.

Conclusions

Folksonomies enable groups to use consistent terminology, for instance, to facilitate searches for items. Folksonomies can develop in crowdsourced systems through

discussion among contributors and users but it requires considerable effort to make the terminology consistent and it is difficult to make decisions visible or to enforce usage. Our examination of the three identified issues identifies the important link between organizational power and technical capability and raises several questions that project designers should consider as they consider how to manage folksonomies.

This short research note has several limitations that present opportunities for continued research. First, we examined only three sites. Other projects may have identified additional approaches to the issues considered. Future research should map different divisions of power and expanding the toolbox of possible interventions from a broader range of projects. Second, as we noticed in Table II, most attention seems to have been given to gardening techniques that intervene when individuals tag and object (upper left corner) or when a folksonomy has emerged (lower right corner). Little research has considered gardening techniques that help the community reach consensus around a single object (lower left corner) or promote the development of coherent personomies (upper right corner). Intervening at these points could lead to more coherent and useful folksonomies later in the process that promote a strong shared language and easy retrieval of relevant objects.

Acknowledgments

References

Al-Khalifa, H. S. and H. C. Davis (2007): 'Exploring the Value of Folksonomies for Creating Semantic Metadata'. *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 3, no. 1, pp. 12–38.

Ames, M. and M. Naaman (2007): 'Why We Tag: Motivations for Annotation in Mobile and Online Media'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. San Jose California USA, pp. 971–980, ACM.

Aurnhammer, M., P. Hanappe, and L. Steels (2006): 'Integrating Collaborative Tagging and Emergent Semantics for Image Retrieval'. In: *15th International World Wide Conference (WWW)*.

Begelman, G., P. Keller, and F. Smadja (2006): 'Automated Tag Clustering: Improving Search and Exploration in the Tag Space'. In: *Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland*. pp. 15–33.

Bonney, R., C. B. Cooper, J. Dickinson, S. Kelling, T. Phillips, K. V. Rosenberg, and J. Shirk (2009): 'Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy'. *BioScience*, vol. 59, no. 11, pp. 977–984.

Bullard, J. (2019): 'Curated Folksonomies: Three Implementations of Structure through Human Judgment'. *Knowledge Organization*, vol. 45, no. 8, pp. 643–652.

Chun, S. and M. Jenkins (2005): 'Cataloguing by Crowd; A Proposal for the Development of a Community Cataloguing Tool to Capture Subject Information for Images (A Professional Forum)'. In: *Museums and the Web*. p. 6.

Crowston, K. and I. Fagnot (2018): 'Stages of Motivation for Contributing User-Generated Content: A Theory and Empirical Test'. *International Journal of Human-Computer Studies*, vol. 109, pp. 89–101.

Crowston, K. and E. E. Kammerer (1998): 'Coordination and Collective Mind in Software Requirements Development'. *IBM Systems Journal*, vol. 37, no. 2, pp. 227–245.

Dieckman, C. (2022): 'Comparisons of Aboutness of OCLC FAST Headings and IMDb "Plot Keywords" for Documentary Films'. *Journal of Library Metadata*, vol. 22, no. 1-2, pp. 135–148.

Doan, A., R. Ramakrishnan, and A. Y. Halevy (2011): 'Crowdsourcing Systems on the World-Wide Web'. *Communications of the ACM*, vol. 54, no. 4, pp. 86–96.

Dougherty, D. (1992): 'Interpretive Barriers to Successful Product Innovation in Large Firms'. *Organization science*, vol. 3, no. 2, pp. 179–202.

Feinberg, M. (2006): 'An Examination of Authority in Social Classification Systems'. *Advances in Classification Research Online*, pp. 1–11.

Fortson, L., K. Masters, R. Nichol, K. Borne, E. Edmondson, C. Lintott, J. Raddick, K. Schawinski, and J. Wallin (2012): 'Galaxy Zoo: Morphological Classification and Citizen Science'. In: M. J. Way, J. D. Scargle, K. M. Ali, and A. N. Srivastava (eds.): *Advances in Machine Learning and Data Mining for Astronomy*. CRC Press.

Guy, M. and E. Tonkin (2006): 'Tidying up Tags'. *D-lib Magazine*, vol. 12, no. 1, pp. 1082–9873.

Hine, C. (2008): *Virtual Ethnography*. SAGE Publications Ltd.

Hjørland, B. (2008): 'What is Knowledge Organization (KO)?'. *Knowledge Organization*, vol. 35, no. 2-3, pp. 86–101.

Howe, J. (2010): 'Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business'. *American Journal of Health-System Pharmacy*, vol. 67, no. 18, pp. 1565–1566.

Jackson, C., K. Crowston, C. Østerlund, and M. Harandi (2018): 'Folksonomies to Support Coordination and Coordination of Folksonomies'. *Computer Supported Cooperative Work (CSCW)*, vol. 27, pp. 647–678.

Lam, W. S. E. (2008): 'Language Socialization in Online Communities'. In: N. H. Hornberger (ed.): *Encyclopedia of Language and Education*. Boston, MA: Springer US, pp. 2859–2869.

Landbeck, C. (2007): 'Folksonomies: Trouble in Paradise: Conflict Management and Resolution in Social Classification Environments'. *Bulletin of the American Society for Information Science and Technology*, vol. 34, no. 1, pp. 16–20.

Lee, D. H. and T. Schleyer (2012): 'Social Tagging Is No Substitute for Controlled Indexing: A Comparison of Medical Subject Headings and CITEULIKE Tags Assigned to 231,388 Papers'. *Journal of the American Society for Information Science and Technology*, vol. 63, no. 9, pp. 1747–1757.

Lu, C., J.-r. Park, and X. Hu (2010): 'User Tags Versus Expert-assigned Subject Terms: A Comparison of LibraryThing Tags and Library of Congress Subject Headings'. *Journal of Information Science*, vol. 36, no. 6, pp. 763–779.

Macgregor, G. and E. McCulloch (2006): 'Collaborative Tagging as a Knowledge Organisation and Resource Discovery Tool'. *Library Review*, vol. 55, no. 5, pp. 291–300.

Mao, M., S. Chen, F. Zhang, J. Han, and Q. Xiao (2021): 'Hybrid Ecommerce Recommendation Model Incorporating Product Taxonomy and Folksonomy'. *Knowledge-Based Systems*, vol. 214, no. Journal Article, pp. 106720.

Miller, P. (2005): 'Web 2.0: Building the New Library'. *Ariadne*, vol. 45, no. 30, pp. 10.

Mocnik, F.-B., A. Zipf, and M. Raifer (2017): 'The OpenStreetMap Folksonomy and Its Evolution'. *Geo-spatial Information Science*, vol. 20, no. 3, pp. 219–230.

Munk, T. B. and K. Mørk (2007): 'Folksonomy, The Power Law & the Significance of the Least Effort'. *Knowlege Organization*, vol. 34, no. 1, pp. 16–33.

Noruzi, A. (2006): 'Folksonomies: (Un)Controlled Vocabulary?'. *Knowledge Organization*, vol. 33, no. 4, pp. 199–203.

O'Leary, D. E. (2015): 'Crowdsourcing Tags in Accounting and Finance: Review, Analysis, and Emerging Issues'. *Journal of Emerging Technologies in Accounting*, vol. 12, no. 1, pp. 93–115.

Østerlund, C., K. Crowston, and C. Jackson (2020): 'Building an Apparatus: Refractive, Reflective, and Diffractive Readings of Trace Data'. *Journal of the Association for Information Systems*, vol. 21, no. 1, pp. 10.

Peters, I. and K. Weller (2008): 'Tag Gardening for Folksonomy Enrichment and Maintenance'. *Webology*, vol. 5, no. 3, pp. 1–18.

Price, L. (2019): 'Fandom, Folksonomies and Creativity: The Case of the Archive of Our Own'. *The Human Position in an Artificial World: : Creativity, Ethics and AI in Knowledge Organization*, pp. 11–37.

Rafferty, P. and R. Hidderley (2007): 'Flickr and Democratic Indexing: Dialogic Approaches to Indexing'. In: *Aslib Proceedings*, Vol. 59. pp. 397–410.

Razikin, K., D. H. Goh, A. Y. Chua, and C. S. Lee (2011): 'Social Tags for Resource Discovery: A Comparison between Machine Learning and User-centric Approaches'. *Journal of Information Science*, vol. 37, no. 4, pp. 391–404.

Smith, C. A. and P. J. Wicks (2008): 'PatientsLikeMe: Consumer Health Vocabulary as a Folksonomy'. In: *AMIA Annual Symposium Proceedings*, Vol. 2008. p. 682, American Medical Informatics Association.

Spiteri, L. F. (2007): 'The Structure and Form of Folksonomy Tags: The Road to the Public Library Catalog'. *Information Technology and Libraries*, vol. 26, no. 3, pp. 13–25.

Stewart, B. and K. D. Kendrick (2019): '"Hard to Find": Information Barriers among LGBT College Students'. *Aslib Journal of Information Management*, vol. 71, no. 5, pp. 601–617.

Syn, S. Y. and M. B. Spring (2013): 'Finding Subject Terms for Classificatory Metadata from User-generated Social Tags'. *Journal of the American Society for Information Science and Technology*, vol. 64, no. 5, pp. 964–980.

Trant, J. (2009): 'Studying Social Tagging and Folksonomy: A Review and Framework'. *Journal of Digital Information*, vol. 10, no. 1, pp. 1–44.

Tsui, E., W. M. Wang, C. F. Cheung, and A. S. Lau (2010): 'A Concept–Relationship Acquisition and Inference Approach for Hierarchical Taxonomy Construction from Tags'. *Information Processing & Management*, vol. 46, no. 1, pp. 44–57.

Walker, J. (2005): 'Feral Hypertext: When Hypertext Literature Escapes Control'. In: *Proceedings of the Sixteenth ACM Conference on Hypertext and Hypermedia*. Salzburg Austria, pp. 46–53, ACM.

Yi, K. and L. M. Chan (2009): 'Linking Folksonomy to Library of Congress Subject Headings: An Exploratory Study'. *Journal of Documentation*, vol. 65, no. 6, pp. 872–900.

Zevin, M., C. B. Jackson, Z. Doctor, Y. Wu, C. Østerlund, L. C. Johnson, C. P. L. Berry, K. Crowston, S. B. Coughlin, V. Kalogera, S. Banagiri, D. Davis, J. Glanzer, R. Hao, A. K. Katsaggelos, O. Patane, J. Sanchez, J. Smith, S. Soni, L. Trouille, M. Walker, I. Aerith, W. Domainko, V.-G. Baranowski, G. Niklasch, and B. Téglás (2024): 'Gravity Spy: Lessons Learned And A Path Forward'. *The European Physical Journal Plus*, vol. 139, no. 1, pp. 100.

Zheng, T., P. Seetharaman, and B. Pardo (2016): 'SocialFX: Studying a Crowdsourced Folksonomy of Audio Effects Terms'. In: *Proceedings of the 24th ACM International Conference on Multimedia*. Amsterdam The Netherlands, pp. 182–186, ACM.

Zhitomirsky-Geffet, M., B. H. Kwaśnik, J. Bullard, L. Hajibayova, J. Hamari, and T. Bowman (2016): 'Crowdsourcing Approaches for Knowledge Organization Systems: Crowd Collaboration or Crowd Work?'. *Proceedings of the Association for Information Science and Technology*, vol. 53, no. 1, pp. 1–6.

Zhou, K., C. Yang, L. Li, C. Miao, L. Song, P. Jiang, and J. Su (2023): 'A Folksonomy-Based Collaborative Filtering Method for Crowdsourcing Knowledge-Sharing Communities'. *Kybernetes*, vol. 52, no. 1, pp. 328–343.