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

Carsten Østerlund[1], Corey Jackson[2], Mahboobeh Harandi[1] and Kevin Crowston[1]

[1] School of Information Studies, Syracuse University, Syracuse NY USA
[2] i-School, University of California, Berkeley, Berkeley CA USA

costerlu@syr.edu, coreybjackson@berkeley.edu, mharandi@syr.edu, crowston@syr.edu

Abstract

Members of highly distributed groups benefit from developing a shared language (i.e., specialized terminology to describe their shared work and work situations) to coordinate their activities. 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 inherent tensions associated with the development of folksonomies in crowdsourcing. First, different users of the language may have different needs. In particular, there might be tension if people labelling objects are not the same as those using these labels to search for content. Second, projects need to decide when in the process of language development, they want to intervene to maintain a balance between a stable ontology and the ability of the project to accommodate ongoing changes. Third, who gets to decide on adopted terminology stands as an important problem, and crowdsourcing projects need to decide on a division of power. We illustrate these considerations by reviewing several projects with an emphasis on the Archives of Our Own, the citizen science project Gravity Spy, 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 & Kammerer, 1998; Malone & Crowston, 1994). Without a common interpretive schema, individuals with different backgrounds may interpret situations and tasks differently, hampering collaboration and communication (Dougherty, 1992). For instance, a crowdsourcing project like Snapshot Serengeti, where volunteers label the species of animals seen in pictures from African camera traps, would not work if every volunteer had their own definition of animal species. In this case, the shared language to be used is set by external authorities (i.e., the definition of species given by a zoologist to a citizen science volunteer). In many other cases though, a shared language needs to emerge from the community's work (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 a shared language and interpretive schemas.

To understand the bottom-up development and use of shared language in highly distributed work arrangements, we explore 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 (2008),

who defined it as the outsourcing of functions once performed by employees to an undefined (and generally extensive) network through an open call. Folksonomy refers to user-generated classification and information retrieval methods where 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 (Bullard 2019; Noruzi 2006; Trant 2009), in order to facilitate retrieval and use of that content. 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.

Studying the role of folksonomies in crowdsourcing allows us to explore three tensions in the development of a language in highly distributed groups. First, in distributed work settings, different workers may have different needs for the shared language. In the setting of folksonomies for crowd work, we see specifically a tension between the needs of users who label content to classify it and the needs of those who would use the labels to search for content. A second related tension has to do with temporality, and when in the work process, shared knowledge plays a role. Does it make more sense to seek the development of shared terminology at the time of labelling an object or later in the process where one has compiled a collection of labeled objects? 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, such changes can be disruptive to the use of folksonomy terms in work, where the users may prefer some degree of stability. A third 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.

The goal of this paper is to review different strategies for the development of folksonomies to understand better what issues become important when promoting the development of a shard language in crowdsourcing settings.

Literature review

To set the stage, we start by briefly reviewing the literature on crowdsourcing and folksonomies.

Crowdsourcing

Crowdsourcing is defined by Howe (2008) as, "the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call.” It is a model for collaboration that uses information and communication technologies (ICTs) and human efforts to solve complex individual, organizational, and societal 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, Ramakrishnan, & Halevy, 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 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 captured by the Hubble Space telescope and report on their morphological and evolutionary features (Fortson et al., 2011). As another example, crowdsourcing has been used to solicit the help

of experts to improve on existing technologies. In 2006, Netflix hosted an open competition where people from around the world worked to improve the Netflix movie recommendation algorithm.

Employing crowds to address complex challenges has been quite successful. In Galaxy Zoo, in addition to creating a useful database of galaxies (the main purpose of the project), volunteers have made discoveries that advance astronomy research (Cardamone et al., 2009; Lintott et al., 2008, Lintott et al., 2009). The winners of the Netflix Prize improved the performance of the original recommendation algorithm by 8.4 percent (Hallinan & Striphas, 2015). Other examples of successful crowdsourcing projects include the popular online encyclopedia Wikipedia, where people volunteer to write and edit articles, and the deluge of projects on Amazon's Mechanical Turk where requestors pay people to conduct human intelligence tasks (HITs), such as annotating digitizing handwritten notes or making English language translations of non-English texts.

In many crowdsourcing systems, 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, benefits such as improved language quality, conversational grounding, and retention, providing a community with sociability, support, information, and a sense of collective identity are realized (Lam, 2008).

Climate CoLab is an example of crowdsourcing that has a heavy emphasis on the type of collaboration and coordination that necessitates the need for a shared language. Climate CoLab is hosted by MIT's Center for Collective Intelligence and recruits people to collaboratively draw up proposals to address climate change. As of 2020, it has nearly 125,000 members. Developing a proposal means contributors must conduct research, write, and share results with their collaborators. If the proposals are to be taken seriously, collaborators need to be aware of the linguistic norms and practices for presenting plans and communicating in a language that will be accessible to other participants.

While the shared language is a common need, the characteristics that define many crowdsourcing projects make it challenging to achieve. First, the composition of the crowd may influence a project's ability to achieve a shared language. As most crowdsourced projects recruit members through open calls, the resulting "crowd" has diverse educational attainment, expertise, and commitment to the project. 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, 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. For instance, in Wikipedia, contributors can take on responsibilities such as fighting vandalism and crafting policies (Arazy, Ortega, Nov, Yeo, & Balila, 2015). 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.

Finally, there are 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 language. These challenges will be exacerbated if there

are no language training materials with which to learn the language. Finally, projects may lack processes for adjudicating a new language in the project. Jackson et al. (2018) describe this phenomenon in the case of Gravity Spy, where the secondary task of developing new glitch names produced an overabundance of terms, many known only to the individual who first introduced the term. The authors further emphasized the need for computational tools, e.g., machine learning algorithms, to adjudicate divergent language contributions.

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 labelling of objects, collections, and searches in Internet-based platforms. There is a grass-roots element to the production and use of folksonomies where the labelling practices and resulting folksonomies service emerging communal purposes. Second, labels need to serve the needs of both individuals trying to organize their own materials and searchers looking for materials. Third, the set of labels is often open-ended, so the size to which and speed with which folksonomies grow can defy the temporal rhythms associated with traditional knowledge organization systems.

Folksonomies can be contrasted to formal knowledge organization systems (KOS) and computational approaches to information organization (Bullard 2019). The controlled vocabulary associated with a formal KOS is (1) developed, applied, and controlled by experts, (2) highly-structured, and (3) stable [REF]. On the other hand, computational approaches generate search terms algorithmically, which reduces the need for expert labor. These approaches benefit from large corpora on which to run analysis (Zhitomirsky-Geffet et al. 2016).

Enabling online resource users to contribute keywords or labels for personal and shared information organization and retrieval has gained attention in a broad 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 Mai Chan 2009; Lu, Park, and Hu 2010; Stewart and Kendrick 2019). 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, 2006). One also finds a number of scholars interested in folksonomies for geo-spatial mapping (Mocnik et al., 2017), healthcare (Smith and Wicks 2008), multimedia (Zheng et al., 2016), and finance (O'Leary 2015). This broad interest has also led to a proliferation of overlapping terms. 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 & Jenkins, 2005) and ethnoclassification (Walker, 2005).

Findings: Folksonomies in Crowdsourcing

Next, we discuss how folksonomies work in crowdsourcing settings to illuminate the processes by which shared language can emerge and be used in distributed work settings.

Process of development: Folksonomies emerge out of both individual and communal practices that involve the labelling 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 labelling (or tagging) individual items, participants seek to assign them meaning. As they label multiple items, they may seek to be consistent in the labelling so that the labels can serve to connect related objects. 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 labels do not necessarily refer to the content of the objects but can also denote author, origin, data form, work process, or other characteristics of the object salient to the individual, i.e., highly context-dependent or personalized in ways that only mean something to the person adding them to the collection (Macgregor and McCulloch 2006). For instance, it is not uncommon to find self-directed labels, such as "to read."

	Single Objects	Collections of Objects
Individual	Object labelling	Personomy
Consensus	Object-labelling consensus	Folksonomy

Figure 1. Folksonomy creation process

At the collective level (lower left quadrant in Figure 1), multiple users examining the same object may strive to reach an agreement about the most appropriate labels to describe that object. A common use of labels facilitates search and retrieval, where content created or labelled by one user can be found by interested others. Agreement on labelling may come through discussions about a particular object or more directly when one individual mirror the observed practices of others, and so apply the same label. When visible to others, individual labelling practices may serve as a guide for other's practices.

Folksonomies take shape as personomies and collectively-agreed-upon labels accumulate over time. Discussions about these labels may further facilitate consensus by allowing individuals and groups to base their labelling on previously-shared labels. In other words, the arrow between the lower left and right quadrant goes in both directions, as a developed folksonomy provides the basis for group members to coordinate their labelling and improve retrieval of relevant objects in their searches.

Shortcomings: The free-labelling practices feeding the production of folksonomies rarely achieve a coherent and structured ontology. Instead, most folksonomies resemble what Peters and Weller (2008) named "tag gardens," where each label (or tag) in a folksonomy is like a plant growing wild. A few labels may receive much attention, but often many others proliferate (i.e., a long-tail distribution (Munk and Mørk 2007)), yielding an unruly and overgrown garden in which it is hard to identify the important members. Research has revealed further specific shortcomings of folksonomies (Al-Khalifa and Davis 2007; Lee and Schleyer 2012; Aurnhammer et al. 2006; Guy and Tonkin 2006; Spiteri 2007). User-contributed terms may overlap by being spelling variants, abbreviations, initialisms or acronyms, synonyms, compound labels where two words are joined, broader or narrower terms or comparable thesaurus descriptors. Terms may also be ambiguous by using neologisms, slang, and jargon (Spiteri 2007). A person's or group's use of labelling strategies may also change over time (Begelman, Keller and Smadja 2006) or subcultures may form as smaller groups develop vocabularies to serve their particular interests (Hidderley 2007).

Gardening techniques: 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 Figure 2). First, a number of strategies attempt to guide individuals in

their labelling strategies (upper left side of Figure 2). 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). Some sites use word clouds to call attention to popular labels to help guide individuals in their selection process, though this approach risks promoting labelling driven by imitation as opposed to thoughtfulness (Munk and Mørk 2007). Other strategies include computationally-derived suggestions for labels (Razikin et al., 2011), where machine-learning-derived suggestions help users increase the precision and recall of their labels and reduce the effort of identifying such labels (Bullard 2019).

Second, strategies can attempt to shape the personomies that feed many folksonomies (upper right quadrant). Curiously, we find few of these. One could imagine a strategy that allowed individuals to easily review and curate their own collection of labels. Likewise, it might also benefit the integrity of the taxonomy if people could compare their personomies with other personomies or the shared folksonomy. Alternatively, a platform could allow participants to review and even edit each other's personomies.

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 labels other users have applied to particular objects and engage in a conversation about their labelling in discussions of labels in the Talk and Discussion fora.

	Single Objects	**Collection of Objects**
Individual	● Autofill ● Spell check ● Knowledge Organization System (KOS) suggestions ● Computationally derived suggestions	● Personomy

Communal	● Object labelling consensus reaches in Talk and Discussion forums embedded in the labelling process.	● Folksonomy = Raw material
		  ● Computational process:
		  ○ Word Clouds and other visualizations
		  ○ Computational techniques to find clusters, label-networks, readable dictionaries, etc.
		  ● Expert curated
		  ● Community curated by:
		  ○ Everybody in a structured process which maintains provenance label history
		  ○ Participants selected by algorithmic reputation
		  ○ Apply, selected & trained

Figure 2. Gardening techniques and the focus of their interventions

Fourth, 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 engages professional knowledge organization designers to review a folksonomy and revise descriptors (Syn and Spring 2013). Computational techniques can be applied to transform folksonomies into coherent taxonomies and ontologies. Clustering, matching algorithms, machine-readable dictionaries, and label networks have all been explored as possible ways to process folksonomies (Tsui, Wang, Cheung, and Lau 2010). 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).

Contrariwise, on some platforms, the user community implements structure into unruly folksonomies. In her review, Bullard (2018) describes three empirical cases that, in different ways, divide up the work and power to make changes to the folksonomy. In the project *LibraryThing*, users not only develop a database of books but also curate a folksonomy to manage the labels applied to the books. Anyone on the project can edit labels, for instance, by combining synonymous labels, separating wrongly combined labels, or setting up a vote to reach a communal decision on important labels. *LibraryThing* keeps track of label names' provenance so users can go back and 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 labels 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 labels 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 label, terms can be noted as synonyms so that divergent terms get replaced. The system assigns the right to edit the folksonomy of labels to users with higher algorithmic "reputation." These are users that the system has judged as good contributors. This group alone can decide to merge or separate labels (Bullard, 2019: 11). A more selective approach to who gets involved in folksonomy editing can be found on *Archives 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 anointed, these editors have significant autonomy to manage their assigned section of the folksonomy.

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

Promoting the development of folksonomy in a crowdsourcing setting raises three core questions that go to the core of shared language development: (1) How does one weigh the support for those who develop terms (i.e., label objects) versus those who use these terms to find content? This question gets to the tension between individual versus collective needs. (2) When does one intervene in the work process to promote the development of a shared language? (3) How does one allocate resources and authority to make decisions? We will address these in turn. We discuss each aspect in turn. Table 3 summarizes this comparison.

Different needs: Supporting the development and use of a coherent folksonomy must weigh the needs of different user groups. The nature of the work and the goal of the project might lead one to weigh one over the other. In particular, it appears to matter who labels objects versus who uses those labels to retrieve objects. If it is the same person labelling and retrieving, one may seek different gardening techniques as opposed to situations where labelling and retrieving is not done by the same people. Different motivations might drive each situation. Ames and Naaman (2007) find that individuals have four main motivations to label items in social labelling systems, including self-organization (i.e., organizing their individual collections), self-communication (i.e., the desire 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). Each motivational factor may also lead one to promote gardening techniques that intervene at the individual or the communal level (see Figure 2). If the purpose of the labelling is mainly driven by individual needs, one may help individuals promote some consistency in their labelling and in particular at the personomy level. Contrariwise, if the goal is to support the work and

communication of a community, one needs to focus energy on promoting consensus in the labelling situation and developing coherent folksonomies. Many sites may need to promote both individual and collective needs.

Archive of Our Own (AO3) serves as an interesting example. The primary activity of the site faces outward. Authors label and curate a folksonomy to increase the visibility of their user-generated content for interested readers. However, there is no effort to change the labels individuals assign to an object. Labelers do not receive suggestions for how to label. Instead, all the effort is placed on fitting the contribution into the folksonomy in an effort to equate terms to improve retrieval.

Zooniverse, the citizen science platform, takes another approach to user-generated labels. As volunteers classify scientific data using the categories provided by the science team, they can also provide their own labels on the platform's discussion fora. For instance, in the *Zooniverse*-hosted project *Gravity Spy,* users classify glitches in gravitational wave data into 22 classes defined by the science team behind the project. However, on Talk, advanced volunteers label images that they believe do not belong to any of the existing classes defined by scientists. The labels found on Talk mainly serve active volunteers, Science team members rarely engage with those labels unless called upon to do so. The project has not organized any effort to garden labels at the individual or communal level in the discussion boards. As a result, the emerging folksonomy is very unruly and suffers for many of the shortcomings associated with open folksonomies. Yet, if volunteers hope to have a proposed new glitch class added to the existing list of approved glitches, they need to go through an approval process that includes the submission of a document detailing the glitch morphology, labels associated with this glitch and example images. Science team members then review the proposal to investigate the glitch class further before they decide if it can get included in the main classification system.

In contrast to the previous examples, *Flickr,* a photo sharing site, seems to primarily support individuals' labelling and collections. Individuals can upload personal photos, add titles, descriptions, and labels to organize their collections. Individuals can also make their images accessible to a defined group of participants (e.g., family and friends) or make them public, even allowing others to add labels to shared images. However, there is no attempt to have participants reach a consensus on labels for a specific photo or to agree on an overall organization scheme.

When in the process and how is stability maintained? Crowdsourcing projects also face the choice of when to intervene in the process to maintain the stability for the folksonomy. Specifically, a project can promote gardening techniques in the labelling situation (as new contributions are labelled) or later in the process at the collections level.

During labelling, users in *AO3* do not seem to attempt to maintain stability. The system does not nudge content creators to adhere to any labelling conventions. Neither do they intervene in the resulting personomies. 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 labels, the provenance of the terms is maintained.

In *Zooniverse*, including the *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. This being said, the team behind Gravity Spy have discussed if they should promote more structured labelling practices at the point where individuals label an object in Talk or the Discussion board. They have considered

autofill, spell check, and computationally derived suggestions that would help volunteers compare their own labels with existing labels.

In *Flickr*, the system only intervenes when individuals upload images by offering computationally-derived label suggestions. Individuals can choose to accept or decline these suggestions.

Allocation of power: Gardening a folksonomy can not only take place at different times in the process but it can also be trusted to different participants in the community. Assigning the gardening tasks also allocates the power to shape the communal language. For instance, some projects might take a laissez-faire approach where anybody can edit the language, while others may choose a democratic strategy involving voting, and others it up to a small group of experts to make the final decisions on shared terminology.

AO3 has bifurcated the power to label individual objects and folksonomies. Individual users generate and own the content and the labels they apply and personomies they develop are not edited by others. However, at the folksonomy level, a select group of participants is given the power to merge, separate, and hierarchically organize labels to develop a coherent folksonomy that is easy to search. A small group of (~300) volunteers to complete a recruitment and training process before gaining permission to curate. At this point, they are free to manage their assigned section of the folksonomy.

On *Gravity Spy* the data labels by volunteers belong to the science team and they retain the final decision on what terms get included in the approved list of terms. In contrast, one finds little oversight and division of power on the Talk and the Discussion boards. A few moderators selected among active volunteers by science team members monitor the discussion and help answer questions and moderate conflicts. But they hold little power to sanction or shape other participants' labelling behaviors. Each volunteer is free to develop and maintain their own personomy.

In *Flickr,* individuals are the owner of their data and generally are the only ones who have authority to label their photos. They may choose popular labels to make their photos searchable among the community or allow certain people in their social network to add labels to their photos. Consequently, there is no authority nor any specific process to curate the folksonomy and clean up the labels to control and maintain controlled vocabularies.

	Archives of Our Own (AO3)	**Zooniverse**	**Flickr**
Who labels versus Who uses labels (Individual vs collective)	● Labelers and users are not predominantly the same. ● Does not change user-chosen labels ● Only equates terms in retrieval by managing labels at folksonomy level	● Two-step process: a. labelling in Talk and Discussion serves the person doing the labelling and fellow volunteers. b. Gardening the folksonomy helps improve consistency and retrieval by the science team and volunteers.	● Accommodates both contributors and searchers but it mainly seems to support contributors and their personomies. ● Attempts to intervene at the individual object labelling level.
When in the process	● Intervenes late in the process by managing the	● Intervenes late in the process.	● Minor intervention early at the time of

	folksonomy. ● Only equates terms at the folksonomy level for retrieval purposes. ● Maintain provenance of label history.	● Only a few new terms get added to the official classification system. ● Unstable in the short term but seeks stability in the long term.	labelling objects by offering computationally derived label suggestions. Fast-changing, with no need for stability
Power	● Bifurcation of power: a. At the personomy level, users generate and own data and maintain their own labels. b. At the folksonomy level, a small group is recruited and trained before gaining permission to curate folksonomy.	● Science teams generate and own data. ● Scientists have the final say in what is regarded as part of the terminology	● Users generate and own their own data. They retain the power to develop and maintain labels.

Table 3. Central tensions in the development of a shared language

Conclusion

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.

Three main tensions appear to play a central role to understand different crowdsourcing strategies to the development and maintenance of folksonomies. First, the goals of different users of the folksonomy need to be weighed. In particular there might be a tension if people labelling objects are also the same as those using these labels to search for content. In this situation interests might not be aligned among the individual people labelling and the collective need to facilitate receival of shared objects. Some projects appear to solve this problem by maintaining two systems. For instance, the project *Archives of Our Own* allows individuals to develop and maintain their own labels. At the same time, they give a small group the authority to garden the shared folksonomy in ways that facilitate search and retrieval.

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. These could serve as the basis for the curation of folksonomies that draws selectively from all the projects collection of personomies.

Second, 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 labelling or wait for a folksonomy to develop before cleaning up the classification of contributions. These considerations need to take

the needs of newcomers into consideration. New participants might find it helpful with some guidance on how to label objects as they get started. Only later may they develop an appreciation of the freedom to choose their own labels. Newcomers can also be overwhelmed by an unruly folksonomy with many overlapping and hard to interpret terms. The *Gravity Spy* project suffers from such maladies. As the project matures so does the size and complexity of its folksonomy where thousands of volunteers have added their own labels with little coordination. With no explicit guiding newcomers struggle to figure out what terms carry more authoritative power.

Third, who gets to decide on adopted terminology stands as an important problem. As we see from the projects described it is not a simple choice between a top-down or bottom-up approach. Several projects like *A03* 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 labels 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 labor and systematic research is needed to map possible divisions of power in a broader range of projects.

Future research may benefit from expanding the toolbox of possible interventions. As we noticed in Figure 2, most attention seems to have been given to gardening techniques that intervene when individuals label 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.

References

Al-Khalifa, H. S., & Davis, H. C. (2007). Exploring the value of folksonomies for creating semantic metadata. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *3*(1), 12-38.

Ames, M., & Naaman, M. (2007). Why we tag: motivations for annotation in mobile and online media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '07* (pp. 971–980). San Jose, California, USA: ACM Press. https://doi.org/10.1145/1240624.1240772

Arazy, O., Ortega, F., Nov, O., Yeo, L., & Balila, A. (2015). Functional Roles and Career Paths in Wikipedia. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1092–1105). Vancouver, BC, Canada: Association for Computing Machinery. https://doi.org/10.1145/2675133.2675257

Aurnhammer, M., Hanappe, P., & Steels, L. (2006). Integrating Collaborative Tagging and Emergent Semantics for Image Retrieval. *In Proceedings of the World Wide Web Conference*, May 23–26, Edinburgh, UK. http://hdl.handle.net/10261/127833.

Begelman, G., Keller, P., & Smadja, F. (2006, May). Automated tag clustering: Improving search and exploration in the tag space. In *collaborative web tagging workshop at WWW2006, Edinburgh, Scotland* (pp. 15-33).

Bullard, J. (2019). Curated Folksonomies: Three Implementations of Structure through Human Judgment. *Ko Knowledge Organization*, *45*(8), 643-652.

Cardamone, C., Schawinski, K., Sarzi, M., Bamford, S. P., Bennert, N., Urry, C. M., … VandenBerg, J. (2009). Galaxy Zoo Green Peas: discovery of a class of compact extremely star-forming galaxies. *Monthly Notices of the Royal Astronomical Society*, *399*(3), 1191–1205. https://doi.org/10.1111/j.1365-2966.2009.15383.x

Crowston, K., & Kammerer, E. E. (1998). Coordination and collective mind in software requirements development. *IBM Systems Journal*, *37*(2), 227-245.

Doan, A., Ramakrishnan, R., & Halevy, A. Y. (2011, April 1). Crowdsourcing systems on the World-Wide Web. Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1924421.1924442

Guy, M. and Emma T. (2006). Folksonomies: Tidying Up Tags? *D-Lib Magazine* 12(1).

Hallinan, B., & Striphas, T. (2014). Recommended for you: The Netflix Prize and the production of algorithmic culture: *New Media & Society*. https://doi.org/10.1177/1461444814538646

Hidderley, R., & Rafferty, P. (1997). Democratic Indexing: An approach to the retrieval of Fiction. *Information Services & Use*, *17*(2–3), 101–109. https://doi.org/10.3233/ISU-1997-172-304

Howe, J. (2009). *Crowdsourcing: why the power of the crowd is driving the future of business* (1st paperback ed). New York: Three Rivers Press.

Jackson, C., Crowston, K., Østerlund, C., & Harandi, M. (2018). Folksonomies to Support Coordination and Coordination of Folksonomies. *Computer Supported Cooperative Work (CSCW)*, *27*(3), 647–678. https://doi.org/10.1007/s10606-018-9327-z

Lam, W. S. E. (2008). Language Socialization in Online Communities. In N. H. Hornberger (Ed.), *Encyclopedia of Language and Education* (pp. 2859–2869). Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-30424-3_214

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*, *34*(1), 16-20.

Lee, D. H., & Schleyer, T. (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*, *63*(9), 1747–1757. https://doi.org/10.1002/asi.22653

Lintott, C. J., Schawinski, K., Keel, W., van Arkel, H., Bennert, N., Edmondson, E., … Vandenberg, J. (2009). Galaxy Zoo: 'Hanny's Voorwerp', a quasar light echo? *Monthly Notices of the Royal Astronomical Society*, *399*(1), 129–140. https://doi.org/10.1111/j.1365-2966.2009.15299.x

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., ... & Murray, P. (2008). Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, *389*(3), 1179-1189.

Lu, C., Park, J., & Hu, X. (2010). User tags versus expert-assigned subject terms: A comparison of LibraryThing tags and Library of Congress Subject Headings: *Journal of Information Science*. https://doi.org/10.1177/0165551510386173

Macgregor, G. & McCulloch, E. (2006). Collaborative Tagging as a Knowledge Organisation and Resource Discovery Tool. *Library Review* 55(5), 291-300.

Malone, T. & Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Comput. Surv.* 26(1), 87–119. DOI:https://doi.org/10.1145/174666.174668

Miller, P. (2005). Web 2.0: Building the New Library. *Ariadne*, (45). Retrieved from http://www.ariadne.ac.uk/issue/45/miller/

Mocnik, F. B., Zipf, A., & Raifer, M. (2017). The OpenStreetMap folksonomy and its evolution. *Geo-spatial Information Science*, *20*(3), 219-230.

Munk, T.B., & Mørk, K. (2007). "Folksonomy, the Power Law & the Significance of the Least Effort." *Knowledge Organization,* 34(1), 16–33.

Noruzi, A. (2006). "Folksonomies: (Un)Controlled Vocabulary ?" *Knowledge Organization* 33(4), 199–203.

O'Leary, D. E. (2015). Crowdsourcing tags in accounting and finance: Review, analysis, and emerging issues. *Journal of Emerging Technologies in Accounting*, *12*(1), 93-115.

Peters, I., & Weller, K. (2008). "Tag Gardening for Folksonomy Enrichment and Maintenance." *Webology* 5(3).

Smith, C. A., & Wicks, P. J. (2008). PatientsLikeMe: Consumer health vocabulary as a folksonomy. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, *2008*, 682–686.

Spiteri, L. F. (2007). The structure and form of folksonomy tags: The road to the public library catalog. *Information technology and libraries*, *26*(3), 13-25.

Stewart, B., & Kendrick, K. D. (2019). "Hard to find": information barriers among LGBT college students. *Aslib Journal of Information Management*, *71*(5), 601–617. https://doi.org/10.1108/AJIM-02-2019-0040

Syn, S. Y., & Spring, M. B. (2013). Finding subject terms for classificatory metadata from user-generated social tags. *Journal of the American Society for Information Science and Technology*, *64*(5), 964–980. https://doi.org/10.1002/asi.22804

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

Tsui, E., Wang, W. M., Cheung, C. F., & Lau, A. S. M. (2010). A concept–relationship acquisition and inference approach for hierarchical taxonomy construction from tags. *Information Processing & Management*, *46*(1), 44–57. https://doi.org/10.1016/j.ipm.2009.05.009

Walker, J. (2005, September). Feral hypertext: When hypertext literature escapes control. In *Proceedings of the Conference on Hypertext and hypermedia* (pp. 46-53).

Yi, K., & Mai Chan, L. (2009). Linking folksonomy to Library of Congress subject headings: an exploratory study. *Journal of Documentation*, *65*(6), 872–900. https://doi.org/10.1108/00220410910998906

Zheng, T., Seetharaman, P., & Pardo, B. (2016). Socialfx: Studying a crowdsourced folksonomy of audio effects terms. In *Proceedings of the 24th ACM international conference on Multimedia* (pp. 182-186).

Zhitomirsky-Geffet, M., Kwaśnik, B. H., Bullard, J., Hajibayova, L., Hamari, J., & Bowman, T. (2016). Crowdsourcing Approaches for Knowledge Organization Systems: Crowd collaboration or Crowd Work? *Proceedings of the Association for Information Science and Technology*, *53*(1), 1–6. https://doi.org/10.1002/pra2.2016.14505301013