

Information Quality Work Organization in Wikipedia

BESIKI STVILIA¹, MICHAEL B. TWIDALE², LINDA C. SMITH², AND LES GASSER²

¹*College of Information, Florida State University, 269 Louis Shores Building, Tallahassee, FL 32306-2100, USA (bstvilia@fsu.edu)*

²*Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 501 E. Daniel Street, Champaign, IL 61820, USA ({twidale, lcsmith, gasser}@uiuc.edu)*

Abstract.

The classic problem within the information quality (IQ) research and practice community has been the problem of defining IQ. It has been found repeatedly that IQ is context sensitive and cannot be described, measured, and assured with a single model. There is a need for empirical case studies of IQ work in different systems to develop a systematic knowledge that can then inform and guide the construction of context-specific IQ models. This article analyzes the organization of IQ assurance work in a large-scale, open, collaborative encyclopedia—Wikipedia. What is special about Wikipedia as a resource is that the quality discussions and processes are strongly connected to the data itself and are accessible to the general public. This openness makes it particularly easy for researchers to study a particular kind of collaborative work that is highly distributed and that has a particularly substantial focus, not just on error detection, but also on error correction. We believe that the study of those evolving debates and processes and of the IQ assurance model as a whole has useful implications for the improvement of quality in other more conventional databases.

Key words:

Collaborative Quality Control, Collaborative Content Creation, Information Quality, Distributed Collective Practices

1. Introduction

Large-scale, continuously evolving, open collaborative content creation systems such as Wikipedia have become increasingly popular. At the same time, in an attempt to lower the bottom line, many traditional publishers and information-intensive organizations have opened their content creation processes to the general public by adding wikis and blogs to their regular channels of information creation and distribution. We are witnessing the establishment of a dynamic grid of large-scale, open information systems fueled by active participation from the general public in content creation and quality assurance activities. Although providing valuable information services to the users, the new information grid also poses new and significant challenges in many areas of information organization, including information quality (IQ).

These new systems have complex, dynamic workflows that need to react successfully to changes in both their communities and the environment, including identifying the most effective and efficient IQ assurance interventions for different circumstances. Furthermore, the concept of IQ itself is context sensitive (Wang & Strong, 1996). The same information can be judged as being of different quality depending on the context of a particular use and the individual or community value structures for quality. Hence, no one fixed model of IQ assurance can be applied for all these systems. There is a need for empirical studies of existing IQ assurance models, with a goal to develop a knowledge base of conceptual models of IQ, taxonomies of quality problems and activities, metrics, trade-offs, strategies, policies, and references sources. The knowledge base can then be reused for constructing context-specific IQ assurance models faster, cheaper, and with less effort. The English Wikipedia, with its large-scale, complex, and collaborative information workflow, is a particularly interesting case and offers excellent empirical data on IQ.

Although collaborative knowledge creation and organization have been in practice since biblical times, with scribes transcribing and at the same time often editing, updating, interpreting, or reinterpreting original texts (McArthur 1986), open access, large-scale, public collaborative content creation projects are relatively recent phenomena. The collaborative participation in the creation of the Oxford English Dictionary (Winchester, 2003) is the best-known example in a precomputer age. Thousands of volunteers submitted examples of the earliest usage from published literature, which were collected, verified, and incorporated into the mammoth project. The postal service was the major remote collaboration medium. As with more modern, widely distributed collaborative activities, including much open source software development, the Oxford English Dictionary project was easy to

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008).

Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

break down into separate activities. However, it required central organization and authority to assemble, assess, and reconstitute the final product from the contributions, many of which were inevitably incorrect, obsolete, or duplicated. Substantial voluntary participation allowed for the discovery of the first published usage of words, a task that is extremely difficult in terms of discovery of candidates, but where verification of candidates is much easier.

Such activities are potentially greatly facilitated by new Internet-based content management technologies such as wikis. Ward Cunningham developed the first wiki engine and established the first wiki repository in 1995, as well as coining the word *wiki*.¹ The key characteristic of wiki software is that it allows for very low-cost collective content creation by using a regular Web browser and simple markup language. These features make wiki software a popular choice for knowledge creation projects, where minimizing overhead in creating new content or editing and accessing previously existing content are the priorities. One such project has been Wikipedia, the world's largest wiki and online encyclopedia, established in 2001.¹ Wikipedia is a community-based encyclopedia that has seen a huge growth in both size and public popularity. As of September 2007, the English Wikipedia alone included more than 2 million articles and more than 10 million pages in total. In addition to the English language encyclopedia, Wikipedia has encyclopedia projects in more than 100 languages, and since March 2006 has remained within the top 20 highly used Web sites in the world.² As a volunteer project, Wikipedia needs substantial active participation and contribution from the general public to grow and improve. Therefore, it allows any user with Web access to start a new article or modify the existing ones with the least possible effort, and it commits and renders the contributions immediately on the user's screen.

To an outsider, the idea of Wikipedia sounds so bizarre and anarchic that it is entirely understandable that there is much skepticism about its quality: How can the entries be any good? Can they be trusted? And won't they just get destroyed or maliciously modified by vandals? These are very reasonable questions, and have led to a number of studies to investigate them (Emigh & Herring, 2005; Giles, 2005). Attention has been given to the quality of the articles, particularly in comparison with equivalents in conventional encyclopedias. These studies appear to indicate that the quality is reasonably good or, at the least, nowhere near as bad as might be expected from such a mechanism. In this paper, we do not try to assess that relative quality directly, but rather try to understand the

¹ <http://c2.com/cgi/wiki?WikiHistory>

² http://meta.wikimedia.org/wiki/Wikipedia.org_is_more_popular_than...

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008). Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

collaborative quality control processes used in Wikipedia and how they manage to achieve the kinds of success that others have noted. This paper is part of a larger body of work looking at IQ and IQ assurance models with the goal of developing an IQ knowledge base, which then can be used for constructing context-specific IQ models in a systematic and less expensive way.

The Wikipedia community has created a series of workflow processes that enable the use of powerful mechanisms for improving IQ while using relatively simple collaborative technology. It is certainly not the case that the quality control mechanisms were always substantially in place and have merely been improved somewhat over time. Indeed, unlike many other collaborative activities, there was no pre-existing noncomputer process to serve as a template for computerization. We believe that by understanding the IQ processes of the special case of Wikipedia, we can obtain insights that can be applied to other databases, including commercial ones that might exploit some features or variants of collaborative IQ improvement activities by users of those databases.

Wikipedia raises many questions in common with open source software: (1) Why do people bother to contribute? (2) How “good” is the resultant product (or product-in-time, given constant evolution)? (3) Why do people trust it and use it? (4) Why does the project not just disintegrate into anarchy? (5) How is the project organized, and how do the processes change over time? (see Benkler, 2002; Crowston & Scozzi, 2004; Lerner & Tirole, 2004; and Scacchi, 2005, for reviews of these issues in the case of Free and Open Source Software [FOSS]). FOSS has been in existence since the 1970s with the GNU project and the Free Software Foundation. Large-scale, open access collaborative content creation projects such as Wikipedia (the world’s largest online encyclopedia), however, are a relatively new phenomenon. The low barrier of entry and, consequently, the large size and extreme diversity of its editorial community sharply distinguish Wikipedia from any FOSS project.

This study builds on our previous studies of IQ problems, trade-offs, and measurement in Wikipedia (Stvilia, Twidale, Gasser, & Smith 2005a; Stvilia, Twidale, Smith, & Gasser, 2005b) and focuses on the overall IQ work organization. It explores how quality issues are discussed by the Wikipedia community, how quality assurance processes evolve, and how, by an analysis of this quality and the creation processes used, one can begin to understand why the quality is better than might be expected.

2. Related Literature

Neither the idea of “wikification” of encyclopedia content nor its construction process is new. Before the existence of the Web, when discussing the possible impacts of hypertext technologies on the encyclopedia genre, Smith (1989) predicted that in electronic hypertext-based encyclopedias, article sequence would not be linear and multiple paths would be provided, author and reader roles would be blurred and author contributions would be augmented by reader annotations, and article bibliographies would be partially replaced by direct hyperlinks to the source documents. What is new in Wikipedia, however, are the low barriers to participation; the sheer size, speed, and geographical distribution of the knowledge construction process; and the ease of accessing this process, all enabled by wiki technology and the Web.

A number of studies have looked at the quality of Wikipedia from different perspectives. Lih (2004) studied Wikipedia content construction and use processes from the perspective of participatory journalism. Viegas, Wattenberg, and Dave (2004) developed a tool (History Flow) to visualize Wikipedia content evolution by using article version histories. Emigh and Herring (2005) compared two community-based encyclopedias (Wikipedia and Everything2) to the Columbia Encyclopedia on the formality of language used. Bryant, Forte, and Bruckman (2005) interviewed nine editors to identify how the sociotechnical structures mediate and shape user activity in Wikipedia, and how the user’s motivation and roles evolve over time. More recently, Pentzold and Seindenglanz (2006), using Foucault’s discourse theory, proposed a framework for analyzing Wikipedia content construction processes and demonstrated it on a sample article. Viegas, Wattenberg, Kriss, and van Ham (2007) revisited their earlier study and analyzed how different structural variables of the English Wikipedia have evolved over time. Similar to this and to our earlier studies (Stvilia et al., 2005a, 2005b), they found that the work coordination infrastructure of Wikipedia has been growing and becoming increasingly sophisticated. Wilkinson and Huberman (2007) validated the correlation between the number of edits and the quality of Wikipedia articles by controlling for the variables of article age and popularity.

Although sharing many features of open participation with Open Source Software (OSS) projects (e.g. Sack et al., 2006), Wikipedia has some important differences. Participants do not need to be as technologically sophisticated as the OSS hacker-programmers. Furthermore, participation is immediate—anyone can switch at any time from being a Wikipedia user reading an article to an editor changing that same article. Individual articles have a relationship with

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008). Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

the articles to which they are linked but, to a large extent, each article can be worked on separately by individuals or subgroups, allowing a far finer granularity of participation than is possible in most software development. In addition, the nature of wiki software enables the tight integration of several people working together on editing the same article, but equally enables conflicts to arise, sometimes leading to edit wars.

This paper does not provide a full exploration of all the work organization aspects of Wikipedia. There are clearly major issues of cooperation and conflict, coordination mechanisms, awareness, articulation, workflows, and evolving processes, all mediated, to a large extent, through a single piece of software and used by a very large heterogeneous community. We focus on the evolution of IQ processes and how they work. However, it is worth noting in passing that Wikipedia does successfully manage to address the factors that are important in ensuring success, or at least in avoiding failure. These include the common artifact criteria of predictability and double-level language (Pekkola, 2003; Robinson, 1993) and Grudin's (1994) criteria of the importance of addressing disparities in work and benefits, and processes for addressing issues of critical mass, exception handling, unobtrusive accessibility, and smooth adoption. Additionally, there are many mechanisms allowing the adaptation and evolution of work processes, both for local exceptions and for larger scale evolution of the work processes themselves, by various mechanisms supporting articulation work (Schmidt & Bannon, 1993) and awareness. In particular, the wiki features of a history mechanism, notification features, and the tight integration of a discussion page with each entry, as well as other spaces for setting policy and norms (and their associated history and discussion pages) play a central role in facilitating these core factors.

There is a well-developed body of research on various aspects of IQ in management science and the database world. Most relevant to this research is the study by Strong, Lee, and Wang (1997). Their study identified a set of organizational IQ problem types that may arise because of aggregating information created in multiple contexts to support a particular task, or from using information created in one context in a different context. However, Strong et al. do not address IQ problems caused by *Many to Many* mappings, that is, when information created in many different contexts has to support the needs and IQ requirements of many different activities and perspectives at the same time. These kinds of situations require constant negotiation, compromises, or consensus building, similar to what we observe in Wikipedia.

3. Research Questions and Design

Systematic empirical studies of information work organization models are essential for testing and modifying the existing knowledge, developing new knowledge, or both, which can then be reused for developing new context-specific models of work organization easily and inexpensively. In this empirical study, we examined the information work in Wikipedia to identify the IQ assurance model and understand how and why it worked. In particular, we attempted to answer the following research questions: (1) How does the Wikipedia community understand quality—what are the quality criteria or what is the quality problem structure? (2) What are some of the IQ assurance processes present, including tasks, roles, social arrangements, and strategies used? (3) How did these quality assurance processes and this understanding of quality evolve? (4) What are the quality dynamics of different types of information objects, and can they be modeled? (5) What are some of the motivations that make people contribute? (7) What are some of the IQ intervention strategies used, and in what circumstances?

The Wikipedia community and work organization have evolved around distinct workflow processes. Ultimately, Wikipedia is a production lattice (Kling & Scacchi, 1982). A web of multiple sociotechnical interaction networks have emerged around a shared resource—Wikipedia’s infrastructure and collection. To understand these complex relationships and structures, and how they change and align with the changing context and objectives, it is important to analyze the dynamics of the collection- or community-level structural variables, as well as individual instances of article creation and IQ maintenance processes, including local interactions, relations, strategies used, arrangements made, and roles played.

We used the case study method (Yin, 1988) with a number of qualitative and quantitative techniques to gain an understanding of the IQ variable structure, dynamics, and organizational issues of IQ assurance in Wikipedia. The data for the study were collected and analyzed between March 2005 and September 2007. In particular, the study looked at collection-level variables, such as the ratios of articles and support pages and the ratios of different user or editor groups and their edit shares, to identify Wikipedia work patterns and dynamics at the collection level. At the same time, in a more in-depth study using the grounded approach (Glaser & Strauss, 1967), we examined the process logs and discussion, policy, and user pages to identify both the formal and informal (active) components of IQ assurance work: the IQ problem structure, IQ assurance or degradation activities and actions, the different roles played, and the patterns of interactions among those roles. The findings of the exploratory structural analysis were

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008).

Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

then combined with the findings of the grounded analysis of the Wikipedia process artifacts to identify the evolving IQ assurance model and understand how and why it works better than expected.

For the collection-level macro analysis, we used the March 9, 2005, November 30, 2006, and September 8, 2007 dumps (copies) of the English Wikipedia database. The dumps contained 500,623, 1,774,062, and 2,234,170 articles (after removing redirects to different pages), respectively. From each dump, samples of 1,000 articles were randomly selected. To gain a better understanding of the IQ assessment model used by the Wikipedia community, we also collected the Featured Article sets, that is, articles considered as Wikipedia's best and ones that had been featured³ on its front page by the time the dumps were made (236, 715, and 822 articles).⁴ Furthermore, because the Wikipedia IQ assurance work is not restricted to editing articles and includes developing and maintaining a complex network of support artifacts (policies, projects, categories, templates, software modules, etc.), we generated random samples of 1,000 pages from the projects, categories, and template collections (namespaces). We also automatically harvested ("screen scraped") the edit histories and extracted from the dumps discussion pages for all of the articles and support pages in the samples.

In the grounded analysis, we manually examined small samples of process support artifacts, such as user home pages, vote logs, policy pages, and discussion pages. In particular, two subsets, each of 30 discussion pages, were randomly selected from the Random and Featured Article samples of the March 9, 2005 dump, and a content analysis (Bailey, 1994) was applied to look for IQ problem types and IQ assurance patterns. In addition, to identify the community's active model for IQ assessment, we content-analyzed the logs of 120 Featured Article Removal Candidate (FARC) votes from July 2004 to May 2005.

To identify the motives and incentives that might drive editors to make contributions, we randomly selected and content-analyzed 100 profiles of user pages from the pool of registered Wikipedia users who edited any of the articles from the random sample of the November 30, 2006 dump. To qualify for the analysis, the user profile had to have some content in it beyond the Wikipedia user template or stub. It has been suggested in earlier studies (Stvilia et al., 2005b; Voss, 2005), and also emphasized in interviews with the Wikipedia founder Jimmy Wales (2005), that the bulk of information work is carried out by a relatively small group of people, the core of the Wikipedia

³ Some articles have been awarded a Featured Article status but have not actually been featured on the Wikipedia main page.

⁴ http://en.wikipedia.org/wiki/Wikipedia:Featured_articles

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008). Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

community. To learn about the IQ roles, tasks, and strategies of these core members, we analyzed the user home pages (including the discussion pages) of three editors with the highest closeness centrality scores in each of the November 30, 2006, sample networks (Featured, Random, Category, Project, Pattern; Wasserman & Faust, 1994). The vertices of these networks represented the editors who edited at least one page, and were connected by pages as edges.

Depending on the type of data being analyzed, units of the content analysis were discussion threads, vote instances and user statements. Coding was performed by the authors themselves. Coding schemes used in the content analysis included the types of IQ problems, actions, roles and motivations. We started with applying an open coding procedure to the samples. Resultant codes were iteratively clustered, compared and revised against similar typologies found in the literature (Gasser & Stvilia, 2001) to develop operational typologies or classifications (Bailey, 1994). We then used these classifications to recode the samples.

4. Wikipedia IQ Assurance Context

This section identifies some of the main components of the IQ assurance context of Wikipedia, such as work coordination and support artifacts, roles, and processes (see Figures 1, 2). Wikipedia is constructed through a highly dynamic complex interaction among its members, who contribute to different processes and play different roles. To understand the dynamics of these interactions and the sources of quality variance in Wikipedia, it is important to identify and enumerate these processes and roles, as well as the motivations and incentive structure of the member contributions. Fortunately, the footprints of some of these interactions and contributions are logged and can be obtained from user home pages, discussion pages, and edit history logs.

The quality assurance context of Wikipedia consists of at least 3 types of processes: (1) those that evaluate the quality of an article and act on it directly by modifying, deleting, or changing its status; (2) those that evaluate the performance of Wikipedia editors and select quality assurance agents (administrators, automatic scripts [bots]); and (3) those that build and maintain Wikipedia's work coordination artifacts.

Thus, the aforementioned processes may affect Wikipedia's quality directly by modifying its content, as well as indirectly by modifying the Wikipedia workflow and editor capabilities. This section is organized as follows:

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008).

Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

Section 4.1 looks at IQ assurance (IQA) agent selection and management processes; Section 4.2 examines the use of two main types of work coordination artifacts, discussion pages and edit history logs; and Section 4.3 provides an analysis of Wikipedia’s processes of content quality evaluation.

4.1 WIKIPEDIA IQ ASSURANCE AGENT SELECTION AND MANAGEMENT

Identifying and examining the roles of Wikipedia and the processes in which they operate will help us to understand some of the sources of IQ variance in Wikipedia, as well as to identify the IQ assurance model of Wikipedia as a whole. Our analysis of the Wikipedia talk pages, history logs, policies, and user home pages suggested at least four distinct roles in the Wikipedia content construction process: (1) *editors*: agents that contribute or add new content to the article; (2) *IQA agents*: agents that affect article quality (monitor changes made; revert instances of vandalism; enhance the IQ of the article through minor edits; enhance the IQ of the collection through enforcing IQ norms and criteria across the collection; enhance the IQ of the collection by developing support infrastructure in the form of genre-specific templates and style guides; and foster collaboration and maintain order in article editorial groups); (3) *malicious agents*: agents that purposefully degrade the article quality; and (4) *environmental agents*: agents that change the representational IQ of articles through changes in the real-world states. Although mostly degrading IQ, in a few instances environmental agents may enhance the IQ of an article by aligning the real-world state with the information contained in an article.

Wikipedia allows users to self-select which articles and processes to contribute to and what roles to play. We found that most of the contributions were motivated by user interests in specific areas, often shaped by their organizational and ethnic affiliations, hobbies, professional experiences, and beliefs. In quite a few instances, the users identified the general desire of fixing quality problems and helping the community as a motivation for making contributions:

I like to edit randomly if I see something I feel like improving. I like creating new redirects to help people find pages as well as moving pages when appropriate.

Alternatively, contributions can be driven by a desire to gain visibility and “reputation,” even if this is achieved by disrupting article editorial processes, such as by insulting other editors or inserting inflammatory or inaccurate information, or both:

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008).

Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

The X has well over 50 members now, and has carried out numerous trolling activities. A simple Google search will show that we are well-known in the Slashdot and blogging communities.

Removing the editor–user distinctions and allowing individuals to select themselves for tasks without any prior vetting increases the variance of perceptions of what constitutes good IQ. Many of the editors had strong and often conflicting opinions on what constituted a good encyclopedia article, or an encyclopedia as a whole, and how to achieve it, often expressed on their user pages or in IQ negotiations and votes:

I'm probably older than the overwhelming majority of Wikipedians, I actually remember some of the 1950s popular songs and baseball players I've written articles about. And this, of course, colors my feelings about some of the articles. Not only do I write articles about a lot of older topics, but I try to keep coverage of such topics even by others at a high level. And this has led to some disputes with other Wikipedians who don't seem to think a 50-year-old song, or some other such thing, deserves an article.—User Y.

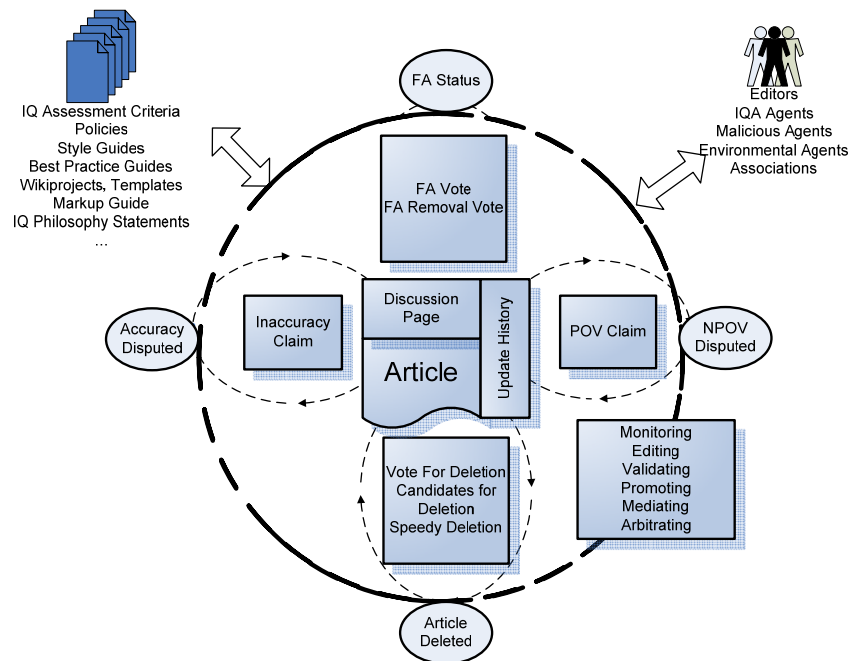


Figure 1: Wikipedia IQ work processes (FA = Featured Article; IQA = information quality assurance; NPOV = neutral point of view; POV = point of view)

The formal organizational structure of Wikipedia is quite well aligned with the roles played by its members. Wikipedia is not a full-blown content management system with finely graded user rights or permission management. However, it distinguishes among three groups of system accounts: (1) *registered users*—identified and tracked by their login name; (2) *anonymous users*—identified and tracked by the Internet Protocol (IP) address they log on from; and (3) *administrators*, including bureaucrats and stewards—the same as registered users but with special system permissions or privileges. Anonymous and registered users may assume different roles at different times.

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008). Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

Registered users can be editors, IQA, or malicious agents. Likewise, anonymous users can act maliciously or can be valuable editors and IQA agents. Although, in theory, an administrator can degrade the quality of an article either intentionally or unintentionally, one mostly expects “good” edits from them to play IQA agent or editor roles.

The organizational structure of Wikipedia is not fixed and constantly evolves together with the actual or active social structure of Wikipedia. Although Wikipedia allows user self-selection of the roles and processes, it also enacts community mechanisms for identifying and distinguishing good editors from malicious ones, and *moderating* their information work resources accordingly.

Editors who are willing to perform quality assurance work are promoted to an administrator or a bureaucrat status based on their performance, knowledge of Wikipedia processes, and most importantly, the needs for additional IQ assurance tools and permissions to make their work even more effective:

It's getting increasingly tedious to manually revert multiple vandalism edits. I know that sooner or later I'll probably have the mop that I need to work efficiently.

Administrators can protect, delete, or undelete Wikipedia pages, including images and their history logs; block and unblock user accounts and IP addresses; use an automatic reversion tool; enforce arbitration decisions; change the wiki software interface wording; hide vandal edits from change logs; move pages; and view hidden pages, including the history of deleted pages. Bureaucrats are administrators and have all the administrator permissions. In addition, bureaucrats can change the name of any user and assign user accounts to automatic processes called bots. They can also promote candidates to adminship.

Nominations or self-nominations for adminship or bureaucratship go through the the Request for Adminship (RfA) and the Request for Bureaucratship (RfB) processes (see Figure 2), by which members of the community, after examining or observing the nominee's past actions, can bring supporting or opposing evidence for the nomination and argue whether endowing the nominee with additional system powers could benefit or harm the project. The general expectation is that a successful nominee is an experienced editor in good standing, with demonstrated knowledge of Wikipedia policies and processes, and is unlikely to misuse additional tools and privileges. After reviewing the submitted evidence, a bureaucrat either promotes the nominee to adminship or declines the request. Although the RfA guide states that decisions should be based on consensus, it also notes that the process is not a ballot, and ultimately the bureaucrat decides what level of consensus is needed to have the

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008).

Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

candidate promoted. This may be intended to avoid collusion and the use of group power to promote a candidate on the basis of kinship, rather than on qualifications:

Prior to mid-2005, RFAs typically did not attract much attention. Since then, it has become quite common for RFAs to attract huge numbers of RFA groupies who all support one another.

The Request for Bureaucrathship is a similar process. However, because bureaucrats are given more power than administrators, promoting them with a higher degree of consensus is expected.

Although administrators and bureaucrats are ranked higher than regular editors in Wikipedia's power structure, they seem to be reluctant to be identified as such and look at themselves as IQ agents ("janitors") who have been given additional tools just for doing cleanup work, and not because of being better editors:

I don't want people to automatically assume I know what I'm talking about by virtue of having admin access—I make mistakes, just like anybody, and I want them to be pointed out to me when I do. I also don't want newbies to think they're not allowed to disagree. I think we should de-emphasize the difference between "admins" and "normal users," not emphasize it. We're all just editors, when it comes down to it.

There is no regular formal process of reviewing an administrator's performance and revoking administrator privileges. Administrators can relinquish or resign from their privileges voluntarily if they leave Wikipedia, find that the privileges may not be needed in their activities anymore, or find it difficult to perform their administrator duties because of misalignment with other administrators:

I've requested desysopping [revoking administrator permissions]. The issue is my judgment. I believe I have very good judgment; others disagree. Therefore my judgment is very poor. A person with poor judgment should not be a sysop.

Administrators who violate Wikipedia's formal policies or informal norms can be demoted by the Arbitration Committee, a semielected body, by using the Request for Arbitration (RfArb) process or directly by the Wikipedia founder Jimmy Wales. Examples of such violations could be the misuse of automatic editing tools, engaging in edit wars with other Wikipedians, violating Wikipedia procedures, or disclosing to third parties the information from hidden pages. Violation claims usually are made by Wikipedia editors or other administrators who may observe or be affected by the administrator's actions. This also means that if unobserved or unreported, administrators may misuse or, alternatively, not use their powers when needed without any reaction or feedback from the community.

There have been attempts to make the management of the administrator group more preemptive. Proposals have been made to establish a formal process for evaluating administrator performance periodically and to ensure that they still garner the community's support and perform the duties for which they requested the tools. The proposals, however, have not gained much support from the community:

People like to point out that we have 200 sysops and that this is not enough. We do and we don't. Many of the officially listed sysops no longer contribute regularly. Some have left Wikipedia altogether, so that, in fact, there is a far smaller number of active sysops who patrol Recent Changes and do what they can to ensure the high quality of our content (and see that the content is provided in an amicable manner).

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008).

Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

I would therefore like to see a situation in which sysops are up for reelection at regular intervals (every three or six months). Let everyone decide who should still be a sysop and who should not. Such a decision would be based on reputation, number of edits, and most of all, trust.

On rare occasions, administrators may radically change their behavior and can qualify as vandals. Alternatively, vandals and trolls can switch roles and become IQ assurance agents:

I find RickK a bit annoying on talk channels, he doesn't seem any worse than User:ClockworkTroll, now an admin . . .

Although at the time of writing, there was no regular formal processes for evaluating and ranking editor performance, there was an informal practice of expressing appreciation and satisfaction with an editor's work by placing an award template⁵ on the editor's home page:

I am awarding you this in acknowledgment of your many proofreads, good edits, and vandal reverts that have benefitted my work and (I must assume) the work of countless others.

Editors and administrators are often assisted by bots (automatic processes) to automate simple, repetitive tasks such as updating templates, greeting or warning users, spell-checking, mining edits for vandalism, and checking images and content for copyright violations. Although any user may run a bot, a bot request needs to go through the Bot for Request for Approval process. The bot needs to be registered, tested, and approved first by the Bots Approvals Group (BAG) and then assigned a user name by a bureaucrat. To remain in the system, the bot needs to show that it is harmless, useful, does not put too great a load on Wikipedia servers, and has been approved and abides by all guidelines, policies, and common practices.

Bots can be very effective IQ assurance tools, but if they malfunction or are used maliciously, bots can also cause significant harm to the project. Administrators are expected to monitor bots closely and block them immediately if their performance does not match the descriptions with which they are registered:

Well the fake bot was simply creating new pages and redirecting the pre-existing ones to them. The number of bots increases. You cannot fight them all manually. The closure of ru-sib drags ridiculously long. We have to push the wikimedia admins to do their job.

The share of information work performed by Wikipedia administrators is astounding, but this is slowly decreasing. Wikipedia posts the list of administrators along with the list of bots. As of March 2005, there were approximately 431 Wikipedia users with administrative privileges and the community employed approximately 50 bots. By November 2006, these numbers had more than doubled (approximately 1,060 and 150, respectively) and by September 2007, the numbers were approximately 1,380 and 372, respectively. The analysis of the random sample edit histories showed that by March 2005, Wikipedia administrators had more than a quarter (approximately 26%) of the total number of edits. Likewise, for the Featured Article set, administrators made approximately 23% of the

⁵ <http://en.wikipedia.org/wiki/Wikipedia:Awards#Barnstars>

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008).

Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

edits. By September 2007, the share of administrator edits was reduced substantially and constituted only approximately 6% of the edits. A similar trend was observed for Featured Articles. The share of administrator edits was reduced to 11%. Interestingly, the share of administrators in the total number of editors for Featured Articles for the same time period was reduced almost 10-fold from 2 to 0.3%. Similar patterns were found for various support pages and constructs, such as templates, which were usually begun by administrators but in which the community had become involved over time, reducing administrator involvement in the maintenance of these artifacts (see Tables 1, 2).

Furthermore, our analysis of edits and votes showed that member participation in Wikipedia’s information work may exhibit a power law pattern (e.g. a few editors vote in almost every IQ vote, whereas many vote in a few or do not vote at all). Based on this statistic and the success of the Wikipedia project, one might suggest that the presence of heavy contributors or power editors is essential to hold the community together and foster the application of common IQ standards and conventions. The majority of Wikipedia editors probably have never read the Wikipedia IQ criteria or style guides. Power editors, on the other hand—who in many cases, but not always, are administrators—are familiar with the IQ policies and norms. One might suggest that those power editors are the ones who follow the standards in their edits and promote and enforce their use in the community.

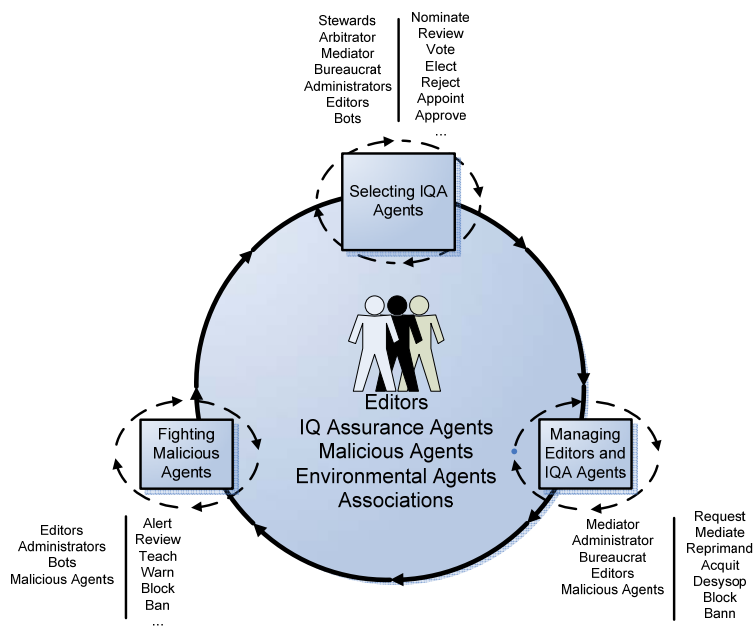


Figure 2: Wikipedia editor management processes

The reduction in administrators’ share of edits and in editors identified by the samples may point to the existence of a robust community that is eager to contribute. At the same time, it may point to a potential scalability problem for Wikipedia with regard to quality. It can be suggested that over time, the administrators may become overpowered by the faster growing general editor population and may not be able to maintain the same level of presence and IQ control. Interestingly, in contrast to the Random samples, the reduction in the administrator edit share for the Featured Article sets was not that high (see Table 1).

Furthermore, as the size of the administrator group has increased, it has become a challenge to communicate information effectively across the administrator group itself and keep its members updated and aligned with Wikipedia’s policies and established practices:

Nothing personal, . . . , but I’m fed up with sysops wandering in, compromising the main page’s integrity (to the embarrassment of the entire community), and pleading ignorance. As the number of admins increases, this problem is only going to get worse, and we shouldn’t have to depend on bots to mitigate the damage.

Measures	March 9, 2005 dump		November 2, 2006 dump			September 8, 2007 dump		
	Random	Featured	Random	Random talk	Featured	Random	Random talk	Featured
Share of administrator edits	0.26	0.23	0.12	0.15	0.18	0.06	0.09	0.11
Share of the administrator group who edited an article in the sample	0.66	0.77	0.53	0.13	0.8	0.31	0.08	0.52
Share of administrators in the number of editors of the sample	0.07	0.02	0.05	0.13	0.005	0.04	0.10	0.003
Share of bot edits	0.05	0.01	0.05	0.01	0.01	0.06	0.37	0.01

Table 1: Distribution of edits in the article samples

Often, these misalignments and differences in editorial philosophies can result in confrontations and edit wars among administrators, which only can be resolved by decisions of the arbitration committee:

Look. I know we have our differences and have spoken many words over the subject. I would like to see this issue resolved but that can’t be done with an edit war, only discussion. If we have to take this to mediation or RFC or arbitration then so be it, but edit warring will not be the answer.

Roles are not the only organizing factors in Wikipedia. Members, while playing the same roles, may hold different perspectives on the effectiveness of specific policies, procedures, or even actions. Like-minded Wikipedia members might form associations to achieve desired outcomes in Wikipedia decision- and policy-making processes. Interestingly, the two most prominent Wikipedia associations, the Association of Inclusionists Wikipedians and the Association of Deletionists Wikipedians, were founded by Wikipedia administrators. Although the Inclusionists

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008). Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

insist that any knowledge is important to at least someone and that all knowledge should be included in Wikipedia, the Deletionists in general are against including obscure, insignificant, or short and poorly written articles.

Deletionists also advocate the adoption and strict application of specific standards and policies on the types and quality of admissible articles. The statements “Wikipedia is not paper” and “Wikipedia is not Google” sum up the two editorial philosophies.

Similar to the Wikipedia roles, the association space and association memberships in Wikipedia are dynamic and informal. The diversity of member IQ preferences and the low cost of forming or switching associations may encourage schism in an existing association or evolution of new groups:

Exclusionism is a form of deletionism, specific to material in articles. . . . The POV [point of view] exclusionist is one who seeks to keep material out of the article for the sake of a POV, whereas a POV deletionist seeks to remove certain Wikipedia articles altogether.

4.2 WORK COORDINATION ARTIFACTS

.It is well understood in Computer Supported Cooperative Work (CSCW) research that effective articulation of work is an essential factor for successful cooperation (Schmidt & Bannon, 1993). Wikipedia successfully uses a number of work coordination artifacts and processes, which we will analyze next.

4.2.1 DISCUSSION PAGES

A discussion page is an auxiliary wiki object that accompanies a Wikipedia article and, as the name indicates, is intended largely for the purposes of communication among the members of the Wikipedia community when constructing and maintaining the article content. Technically, a discussion page (talk) is the same wiki object as an article. Unless locked by Wikipedia administrators, it can be updated by anyone. Updates to the article are logged and can be visualized through a history object. The difference between the article and its discussion page lies only in the role assigned to a discussion page in the Wikipedia infrastructure. It is a coordinative artifact (Schmidt & Simone, 1996) that helps to negotiate and align member perspectives on the content and quality of the article.

Discussion pages are part of Wikipedia’s overall work coordination and support architecture, which also includes topic-specific project management pages called Wikiprojects, style manuals, best practices guides, templates, and other work coordination artifacts. Discussion pages are routinely used by IQA agents such as administrators to

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008).

Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

communicate different kinds of management information, such as providing feedback on quality, giving notices and warnings on the article’s current status, encouraging cross-article communication, and general coordination (see Tables 1, 2).

Furthermore, we found that an article’s discussion page is often used by those outside the article’s editorial group. These outsiders use it to ask the editorial group questions related to the topic of the article, and sometimes even solicit assistance for other Wikipedia articles or projects outside Wikipedia:

Could anyone who is geologically inclined please have a look at Andes. There is currently a warning saying that the section on geology is pre-plate tectonics, so could anyone who knows about such things please have a squiz.

Measures	Project	Project talk	Template	Template talk	Category	Category talk	User	User talk
Share of administrator edits	0.15	0.20	0.14	0.19	0.12	0.07	0.05	0.12
Share of administrators in the number of editors of the sample	0.12	0.22	0.08	0.20	0.11	0.13	0.09	0.13
Share of the administrator group who edited an object in the sample	0.38	0.06	0.15	0.05	0.10	0.02	0.10	0.27
Share of bot edits	0.08	0.01	0.01	0.03	0.17	0.20	0.23	0.07

Table 2: Distribution of edits in the samples of support pages of the September 8, 2007 dump

Traditionally, an encyclopedia article genre is considered to be a concise introduction to a subject in the form of a compressed summary, focusing a reader’s attention on the main points and giving references to outside sources for more in-depth information. Following this convention, Wikipedia articles also may not include detailed explanations or descriptions of the concepts and theories mentioned in the article, sometimes making it difficult for a general reader to fully comprehend. We found that this kind of information was often explicated in the process of negotiation, reflection on edits made, or when answering questions. Accumulated in the discussion pages, such information turned the discussion pages into FAQ-style knowledge bases—complements to information presented in articles and a great resource for regular users:

Your responses to my questions have been very informative. If you get the chance, I hope you’ll consider incorporating some of this material into the article. Although we’re enjoined to be bold in editing, I confess I feel out of my depth here.

Having well-developed work articulation artifacts in the form of discussion pages helps in establishing a sense of community and negotiating a merit-based social order. It helps to establish norms and conventions of communication (Orlikowski & Yates, 1994) and to introduce newcomers to those norms and the subject in general.

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008). Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

We found that the Featured Article set had significantly better developed discussion pages than the Random set. Not only were the median lengths of the discussion pages of the Featured Articles more than 10 times larger, but they were much better organized and more readable based on their Flesch readability scores (Gunning, 1952). The presence of a large and well-organized footprint in the form of a discussion page can be a sign of the existence of a strong interest and a well-organized editorial group around the article. The strong interest, however, may not always be translated into high quality. If there is consensus within the editorial group, a large-sized discussion page can be correlated with the high quality of the article's content. Alternatively, if consensus has not been reached, then a large-sized discussion page can merely be a result of longstanding arguments and disputes.

4.2.2 ARTICLE EDIT HISTORIES

An edit history is a wiki object that contains the log of three-element metadata entries for each edit instance. The edit metadata elements contain the following information: (1) the date and time of the edit; (2) the name of the user who made the edit, or if the user is not logged on with a Wikipedia registered user name, the IP address from which the edit is made; and (3) optionally, a comment provided by the person doing the edit clarifying the purpose of the edit. As a result, the history object can be a source of the following metainformation about the article: (1) age; (2) currency—when the article was last updated; (3) number of times the article has been edited; (4) names or IPs of the article editors; and (5) types of edits, such as reverts (returning the article to an earlier state or version), minor edits, or copyediting.

The last two kinds of information may not be complete and unambiguous. It is not necessary for an individual to be logged on or even registered to make an edit. In addition, the same individual can be registered and make edits with more than one name. Likewise, edits are often made without editors filling out comment fields, or with editors filling them out with misleading information. Because the conventions of commenting edits are not followed consistently, automatic coding of the comments for analysis could not be fully accurate either. Nonetheless, these two elements of history entries can still provide valuable information about the social structure and dynamics of the article's content creation (Stvilia et al., 2005b). We found that editorial groups utilized history logs to coordinate their edits, to identify and fight vandalism, and for auditing in dispute resolution.

4.3 CONTENT QUALITY EVALUATION AND SELECTION PROCESSES

The metaphor used in the Wikipedia quality assurance model is that of “controlled” selection or evolution. In addition to the implicit local quality evaluation and selection carried out by editors every time they modify, delete, or leave an article’s content unchanged, the Wikipedia IQ assurance model has two sets of mechanisms for setting and enforcing community-wide criteria for article quality and controlling the entire IQ profile of the collection:

- Featured Article status assignment and removal processes, which set a target or desired level of quality.
- Article deletion processes, which enforce a minimum level of quality—the lowest critical boundary of quality for an article to remain in the collection (see Figure 1).

It is important to note that a relatively insignificant number of Wikipedia articles have been formally evaluated against the Featured Article Criteria (FAC). The number of Featured Articles, the articles whose IQ has been confirmed as meeting the criteria by community votes, was only 236 (<0.05%) out of the total of 500,623 articles in the Wikipedia collection as of March 2005. By September 8, 2007, this number had grown to 822 articles that had already been featured on the Wikipedia main page. The share of Featured Articles in the overall collection, however, has remained the same, approximately 0.04%. Information on what percentage of new articles or existing articles have been evaluated systematically against the deletion criteria was not available.

We found, however, that the community did not orient itself only toward maintaining the IQ status quo, but also actively sought to improve the overall IQ profile of the collection by deploying various sociotechnical mechanisms of IQ assurance and generating an evolutionary pressure by slowly moving up the target and baseline levels of article quality. We will examine some of these mechanisms in the following subsections.

4.3.1 FEATURED ARTICLE CRITERIA AND PROCESSES

Featured Articles are those declared by the Wikipedia community to be its best. Articles can be nominated as candidates for Featured Article status by individuals or a group. Once nominated, the candidates undergo a peer review process to check whether they meet the Wikipedia FAC. According to the history log of the Featured Article

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008).

Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

directory, the Featured Article process began around April 2002. However, at that time Featured Article candidates did not undergo a peer review process. The directory did not reference any quality assessment criteria except “brilliant prose.” As a result, those early non-peer-reviewed Featured Articles have been referred to ironically by the current Wikipedia community as “brilliant prose” articles:

We think the following Wikipedia pages are pretty good. This is a selected list—since there are thousands of pages on Wikipedia, we couldn't possibly keep track of *all* the brilliant prose here! But if you come across a particularly impressive page, why not add it to the list as your way of saying “Thanks, good job”?⁶

It needs to be noted, however, that at the same time, Wikipedia already had “the Perfect article” page containing a quite well-developed description of what makes a perfect encyclopedia article. It referenced IQ criteria such as clarity, completeness, neutrality, verifiability, and compliance to genre

A widely used general definition for IQ is “fitness for use” (Juran, 1992). A more operational definition of IQ, however, usually includes a set of criteria or dimensions that represent a context-specific understanding of quality. The first ad-hoc formal Featured Article quality assessment guideline was developed by the user Eloquence on March 7, 2004 (personal communication with Marc Pellegrini, the Wikipedia Featured Article Director, August 1, 2005). Eloquence's guidelines included quality criteria—comprehensive, factually accurate, and well written. The guidelines also referenced the Perfect article page as an additional quality guide. The original guidelines underwent multiple modifications and, as of this writing, these now include eight Featured Article quality assessment criteria: (1) comprehensive; (2) accurate and verifiable by including references; (3) stable—not changing often; (4) well written; (5) uncontroversial—using neutral language and not having an ongoing edit war; (6) compliance with Wikipedia standards and project guides; (7) having appropriate images with acceptable copyright status; and (8) having appropriate length, using a summary style and focusing on the main topic. Figure 3 maps the quality assessment dimensions from the printed encyclopedia quality assessment discussion by Crawford (2001) and ones from a generic IQ assessment framework developed earlier (Gasser & Stvilia, 2001) into the FAC.

Although the Wikipedia IQ model lists stable, uncontroversial, and verifiable as important quality dimensions when assessing the quality of a Featured Article candidate, these dimensions do not appear in the Crawford (2001) framework. It could be that in the Crawford framework, they are taken for granted. The content of a printed encyclopedia article is generally fixed until the next update cycle, which is not the case with Wikipedia, where

⁶ http://en.wikipedia.org/w/index.php?title=Wikipedia:Featured_articles&direction=prev&oldid=47610

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008). Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

anyone, including malicious agents, can make edits anytime. Likewise, the FAC do not include the authority and currency dimensions. Although for a multivolume general printed encyclopedia, a yearly revision can be a “Herculean and economically infeasible task” (Crawford, 2001), currency does not seem to be considered a major quality indicator in Wikipedia, where the cost of an update is very low and anyone is allowed to do it. Both the Crawford and the Featured Article IQ assessment models contain accuracy as an IQ criterion.

The Wikipedia community does utilize a reputation mechanism, even though this is not formalized in its IQ assessment criteria. The community is more suspicious of edits made by anonymous or new editors than those made by editors with previously established records of valuable contributions:

Why did you add redlinked personal names to Burr? If an anon had done the same, I would have reverted.

The Wikipedia community also insists on verifiability to enable peer control of the quality of user edits. It places trust not in a single expert author or group, but in the collective knowledge of a large-scale, distributed community, hoping that, when creating encyclopedias, just as for creating OSS, “given enough eyeballs, all bugs are shallow” (Raymond, 1998). Cosley, Fankowski, Kiesler, Terveen, and Riedl (2005) found that peer oversight was as good as expert oversight in maintaining the quality of an information collection. Hence, one of the IQ measures can be the number of “eyeballs” (i.e., the number of distinct editors). Again, this is an indirect measure that happens to be easy to measure. The real number of eyeballs is the number of people reading the article. The number of people bothering to make a change is obviously much smaller and probably more interesting, and may be correlated with the real number of eyeballs. At the same time, one needs to keep in mind that editors reading the article and not making any changes can also be indicators of the good quality of an article, or the article being obscure.

The stability criterion, which assesses how frequently the article undergoes significant changes, and the requirement of having references were not included in the earlier versions of the Featured Article candidate assessment guide and were added only in September 2004. To reduce the IQ variation of Wikipedia content and to make the IQ assessment process more consistent and systematic, the Wikipedia community is rapidly developing sets of style manuals, genre- or topic-specific templates, and information organization and management pages called Wikiprojects. For comparison, the April 2004 version of the FAC contains reference to only 5 guides, whereas the May 2005 version lists 11 guides. The September 2007 version references 17 guides. Hence, the differences between the two IQ assessment models are largely caused by the pragmatics of their immediate social context of use. As

Wikipedia’s social context and available article variability changes, so does its IQ assurance and assessment infrastructure.

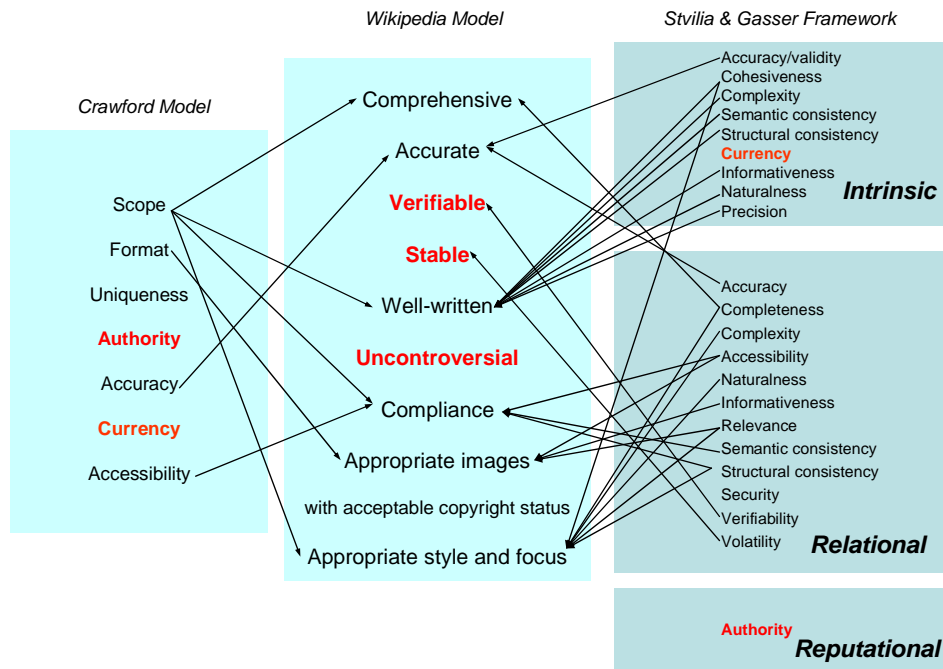


Figure 3: Wikipedia IQ assessment model mapping into the framework of Stvilia and Gasser (Stvilia & Gasser, 2001) and the Crawford model (Crawford, 2001)

As a part of the continuous IQ feedback process, one can also nominate an already Featured Article for removal of the Featured status and have the community vote on it. According to the Wikipedia logs, approximately 120 articles were nominated from July 2004 to May 2005 as candidates for removal, and the community voted out two-thirds of them (80 articles).

Although the Featured Article and FARC process pages emphasize the importance of community consensus when awarding or removing the Featured Article status, like many other Wikipedia processes, they are not completely horizontal but have been shaped strongly by a single person, the Featured Article director, who is also a Wikipedia bureaucrat. However, this may change as Wikipedia grows and the needs and expectations for applying standardized policies and processes across the system clash with local editorial practices and pragmatics.

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008). Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

4.3.2 DELETION CRITERIA AND PROCESSES

The community's IQ reasoning is also encoded in the judgments on the pages that are considered Wikipedia's worst, or not worthy to remain in the collection—deleted articles, stubs, templates, categories, and other miscellaneous pages. According to the Wikipedia Deletion Policy (DP) page, several hundred articles are deleted daily in Wikipedia. The reasons for deletion could be *low quality* (absence of any useful or meaningful content, vandalism, redundant entries), *low criticality* (lack of importance or notability), *copyright violations* or content that is *not encyclopedic* (advertising, self-promotion, original research, or non-Wikipedia-related user pages).

Pages, including articles, can be nominated for deletion by any Wikipedia user. However, they can be deleted or, alternatively, restored only by Wikipedia administrators. Once an article is listed on the Articles for Deletion (AfD) (formerly Votes for Deletion [VfD]) forum, the nomination remains there for a number of days to give the members of the community a chance to vote on it. Decisions on what actions to take (keep, delete, clean up, merge or redirect, or transwiki [move from one wiki to another]) are made on the basis of consensus. A similar procedure applies when someone nominates to undelete a previously deleted article.

To make the weeding process more efficient, however, in certain instances, administrators, guided by the Wikipedia Criteria for Speedy Deletion (CfSD), can delete pages without seeking the consensus of the community on the AfD forum. According to the CfSD, the following kinds of articles can fall under the Speedy Deletion category: (1) very short articles providing little or no context; (2) foreign language articles that already exist in another Wikimedia project; (3) any article whose content consists only of an external link; (4) any article that consists only of attempts to correspond with the person or group named by its title; (5) any article that has been discussed on AfD and successfully transwikified; (6) short articles that serve no purpose but to disparage their subject; and (7) an article about a real person that does not assert that person's importance or significance. In addition, the CfSD include 12 general criteria and separate sets of criteria for different support pages and objects (redirects, media objects, categories, user pages, templates, and portals).

Another semiautomatic deletion process was established in April 2006, the Proposed Deletion (PROD) process. Interestingly, the PROD process does not use a set of deletion criteria different from those of the AfD. The only suggested criterion that differentiates AfD candidates from PROD candidates is “uncontroversial” or “uncontested.”

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008). Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

This seems to be evaluated based on the size of the editorial group for an article and the intensity of its recent activity.

Similar to the AfD, the PROD can be applied to articles, and in some cases to inactive user pages. Similar to AfD articles, PROD articles are listed for 5 days before deletion, but are not discussed or voted on. In addition, in contrast to the CfSD, the PROD process is not used with support pages, such as categories, templates, redirects, or portals. Anyone, including the creator of an article, can object and stop a PROD process simply by removing a PROD tag from the article. Once the tag has been removed, the article cannot be renominated for the PROD, but it can still be listed for the AfD.

Both the CfSD and PROD processes are intended to reduce the cost of IQ assurance by avoiding a discussion or voting step. At the same time, the PROD process serves as a “soft” warning to an article’s editorial group and gives the group a chance to improve the article and avoid deletion. Our analysis of the AfD logs showed that the community often used the AfD process to reinvigorate editorial groups and encourage them to improve the quality of an article. The PROD process may serve the same goal but with a lower cost.

Thus, the criticality or notability of the subject or topic discussed and its appropriateness to the general encyclopedia genre have been two main factors determining whether a particular article needs to be deleted. Quality, on the other hand, unless it is unrecoverable nonsense, has not been strongly emphasized in the DP. It is assumed that the quality of new articles will be low at the beginning.

Interestingly, the earliest version of the Deletion Policy (DP) gave only a brief, informal mention of the deletion process. It did not make any reference to community consensus, placing the burden of decision making on only three administrators who could delete pages on request. In addition, the version did not include *low criticality* or *notability* as a deletion criterion, and the DP criteria overall were informal and less demanding. Finally, it did not distinguish articles from other kinds of Wikipedia objects and referred to all of them as pages. Thus, similar to what we observed for the Featured Article criteria, as the Wikipedia collection and community have been growing, the DP criteria have been evolving as well by expanding and becoming formalized. The AfD, CfSD, and PROD mechanisms have been used to create evolutionary pressure on the collection to reduce variance in the behavior of the editorial community and, ultimately, to reduce variance in the IQ of Wikipedia articles.

4.3.3 ACCURACY AND NPOV DISPUTED

The last two processes from Figure 1, Accuracy Disputed⁷ and NPOV (Neutral Point of View) Disputed⁸ are enacted when the *accuracy* of the article, *neutrality* of the article, or both are disputed and cannot be resolved promptly within the article's editorial group. The person who makes a claim inserts an appropriate tag at the beginning of the article indicating that the quality of the article is being challenged. Note that these tags may serve dual purposes. On the one hand, they indicate to a Wikipedia user that there might be a quality problem with the article. On the other hand, this invites IQA agents, who are often Wikipedia administrators, to come help the editorial group resolve the problem.

5. Wikipedia's IQ Problem Structure and Active Model of IQ Assurance

Quality is ultimately a social construct (Strong et al., 1997). Hence, to evaluate it meaningfully, there must be socially and culturally justified assessment criteria and norms with which to compare it. This implies the existence of some social order or consensus on what constitutes good IQ, which, according to Strauss (1978), is always a negotiated order. Strauss suggests that examining negotiation contexts can help the researcher grasp subtle variations in processes, strategies, and roles that otherwise could go unnoticed, and link them consistently and systematically to the main research topic. In this particular research, we were mostly interested in editor negotiations and discussions as a means for identifying the types of IQ problems, IQ dimensions that were considered important by the community, and trade-offs among those dimensions. These negotiations were combined in our analysis with other processes such as user self-reported descriptions and justifications of actions taken, which helped to uncover Wikipedia's IQ problem structure and the *active* model of IQ assurance. This section starts by identifying the IQ problem structure for the English Wikipedia collection and then examines some of the informal processes and practices of IQ assurance used by the community

⁷ http://en.wikipedia.org/wiki/Category:Accuracy_disputes

⁸ http://en.wikipedia.org/wiki/Category:NPOV_disputes

5.1 IQ PROBLEM STRUCTURE

In the content analysis of 60 discussion pages from the Featured Article and Random sets, we identified 12 types of IQ problems that Wikipedia users pointed to (see Table 3). Problem instances were tallied as many times as they occurred in discussions. It is clear from Table 3 that the distribution of IQ problem types claimed or disputed in the discussion pages was not uniform. We use the word “claim” to refer to a user or editor complaint about an IQ problem that may not be true or agreed upon by the community; IQ assessments are often relative to particular cultural and knowledge structures of a community. If the user is not aligned with those structures, his or her claim that an IQ problem exists may not be shared by the rest of the community and may be rejected.

Quality problem claims often contained information about possible quality improvement or degradation actions, and the sources of quality problems (see Table 3). Collecting and organizing these actions and sources can produce a useful knowledge map, which can be used for organizing and planning IQ assurance activities in similar settings.

Our analysis showed that often when claiming or disputing IQ problems, users were mindful of the existence of trade-offs among various quality dimensions. They sought a balance as a social group among those dimensions through the process of negotiation, logical analysis, and sense making of their own and each other’s actions (Weick, 1995). We identified instances of the following trade-offs from user disputes and self-reported reasoning of the current state or retrospective sense making of the edits they made in articles: (1) Completeness vs. Accessibility; (2) Accuracy vs. Accessibility; (3) Completeness vs. Cohesiveness; (4) Accessibility vs. Complexity; (5) Completeness vs. Consistency (when the same statement may not be consistently interpreted by different cultural or social groups or institutions); (6) Accessibility vs. Consistency; (7) Completeness vs. Complexity; (8) Volatility vs. Accessibility.

Earlier we suggested that a Criticality or Importance measure of an information object to a given community could be orthogonal to its IQ dimensions when doing selection or when making a decision about IQ assurance resource allocation (Gasser & Stvilia, 2003). Examples confirming this proposition were found in the current study. In addition, the community was clearly conscious of constant trade-offs between quality and cost (Stvilia et al., 2005a).

The analysis also identified Accuracy and Completeness as the most frequently encountered IQ problem claims. In examining the Wikipedia IQ model (see Figure 1), we can see that there are separate processes for addressing the Accuracy Disputed and NPOV Disputed claims. This tells us that the Wikipedia IQ assurance model, and the IQ

architecture in particular, was indeed quite well aligned with the contemporary IQ problem structure of the collection (see Figure 1).

Problem types	No. in FA	No. in RA	Caused by	Action taken or suggested
Accessibility	6	3	<ul style="list-style-type: none"> ▪ Language barrier ▪ Poor organization ▪ Policy restrictions imposed by copyrights, Wikipedia internal policies, and automation scripts 	Reorganize, duplicate, remove, translate, split, join, rearrange
Accuracy	54	53	<ul style="list-style-type: none"> ▪ Typing slips ▪ Low language proficiency ▪ Changes in the real-world states ▪ Wording that excludes alternative points of view (POV) ▪ Garbled by software 	Fix, correct, change, remove, revert, remove exhaustive qualifiers, specify, clarify context, update, provide epistemology, verify, explain; resolve contradictions
Authority	2	0	<ul style="list-style-type: none"> ▪ Lack of supporting sources ▪ Lack of academic scrutiny of the sources ▪ Known bias of the source ▪ Unfounded generalization 	Add, replace, remove, reword, qualify
Cohesiveness	1	1	<ul style="list-style-type: none"> ▪ Loss of focus 	Restrict, move
Completeness	49	20	<ul style="list-style-type: none"> ▪ Existence of multiple perspectives ▪ Unbalanced coverage of different perspectives ▪ Lack of detail ▪ Difference between an encyclopedia article genre and the genre from which the text was imported 	Add, specify, disambiguate, include, expound, balance, qualify, clarify, integrate
Complexity	7	8	<ul style="list-style-type: none"> ▪ Low readability ▪ Complex language 	Replace, rewrite, simplify, move, summarize
Consistency	13	12	<ul style="list-style-type: none"> ▪ Using different vocabulary for the same concepts within the article or within the collection ▪ Using different structures and styles for the same type of articles ▪ Nonconformity to the suggested style guides ▪ Differences in culture or language semantics ▪ Conflicting reports of factual information ▪ Contradicting or conflicting with a particular cultural or social norm, convention, or standard 	Reorganize, conform, revert, move, choose the most widely used form, vote
Informativeness	6	4	<ul style="list-style-type: none"> ▪ Content redundancy 	Remove, move, revise, cut down
Naturalness	2	1	<ul style="list-style-type: none"> ▪ Obscure language; text does not flow well 	Edit, rewrite, improve
Relevance	18	16	<ul style="list-style-type: none"> ▪ Adding content that is not relevant or outside the scope of the article 	Revert, move, separate, get rid of, remove
Verifiability	19	12	<ul style="list-style-type: none"> ▪ Lack of references to original sources ▪ Lack of accessibility of original sources 	Add, remove, cite, revert, provide, confirm
Volatility	2	1	<ul style="list-style-type: none"> ▪ Lack of stability caused by edit wars and vandalism 	Avoid, protect

Table 3: IQ problem types, related causal factors, and IQ assurance actions taken or suggested (FA = Featured Articles; RA = Random Articles)

5.2 ACTIVE MODEL OF IQ ASSURANCE

Discussion logs and vote logs can also be used for inferring Wikipedia’s *active* model of IQ assurance. Some parts of the model may be formalized into policies and procedures, whereas others may stay informal but actively used and embedded in the community’s organized activities and actions. For instance, the FAC do not include the Authority or Reputation dimensions. However, Viegas et al. (2004) found in an earlier study, which was

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008). Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

corroborated by the findings of this study as well, that the community does take into consideration the reputation of the editor when making IQ judgments. The records of IQ problem claims and votes may also help in inferring specific trade-offs among different IQ dimensions and the distributions of IQ problem types. Our analysis of the FARC discussions suggested that the odds of Featured Articles losing their status because of Volatility or Stability problems were more than twice as high as the odds occurring because of Authority problems (see Table 4). Similarly, Completeness might be valued significantly higher than Cohesiveness in the Featured Article process.

Although the community consistently pointed to the Featured Article quality criteria when nominating an article for removal or discussing its IQ, the quality criteria themselves kept changing over time. For instance, the requirement that a Featured Article must supply references was added only in September 2004. Consequently, the articles that were well qualified as Featured Articles under the old IQ requirements did not do so well once the requirements changed. The community was willing to compromise and allow the article to retain a Featured Article status if they saw a lively interest in the article topic and members were willing to put some work in the article to make it meet the IQ standards. In 17 instances out of 120 nominations for removal, editors updated articles and addressed some of the criticisms posed in the nominations. In 14 of those 17 cases, the articles were allowed to retain their Featured Article status (see Table 4). Thus, an article on which some IQ improvement work was done had more than a 3-fold greater chance of retaining its Featured Article status compared with an article that had not, 0.82 vs. 0.25:

I added some further reading in lieu of knowing what references were actually used. Further I noticed that the none of the objectors had contacted the original author (still an active wikipediaian) to comment on the issues, so I did so.—User X

Well done—at least these nominations are triggering improvements to the articles, which is clearly a good thing.—User Y

IQ assurance decisions made based on immediate context pragmatics were not rare in the Wikipedia community.

Editors prioritized or optimized their IQ assurance activities based on current events or anniversaries:

We desperately need help with the April fools main page article, theres 4 weeks left and I really doubt we're going to get the stubs that have been suggested up to featured status in time. Have you got any ideas of good or A grade articles we could use? I'm more than happy to do a solo mission on it because interest seems to have been lost—just getting a stub up there isn't going to happen.

In some instances, the community showed an awareness of the risks that open content and vandalism might pose to Wikipedia users, as well as the benefits that the same openness and exposure to diverse perspectives and knowledge could bring to the article:

Many people access wikipedia without fully understanding what wikipedia is all about (because they arrived via a google search or similar) and so they see the vandalised text and believe that to be accurate.

Remember that featured articles that are displayed on the main page are still works-in-process. It is important that new users can edit them.

Identifying effective IQ assurance strategies and resource allocation was also a concern. For instance, members were well aware that each instance of a FARC nomination and vote consumed the community’s resources (time), which instead could be spent on fixing the IQ problem itself. Editors often voiced suggestions to better align IQ assurance activity structures with individual IQ problem types and to optimize IQ assurance resource allocation among different actions of the activity:

These little objections about lead sections, extra references, images, etc., should stop being used as de-feature fodder. If the prose and scope of an article is strong (as it is here), go in and do the modernizing touchup changes yourself. Half the time it will take you barely more effort than starting and defending a nom [nomination] on this page.

Likewise, we found that editors were conscious of the effectiveness of communication and the use of different coordination artifacts when making changes in articles. For instance, voting or polling was used quite often when editors intended to modify the content of a Featured Article, but this was not used for regular articles. The implementation of three different article deletion processes seemed to be intended to reduce the cost of deletion and to use voting only when necessary.

IQ problem types	Status retained	Status lost	Total	Retention probability	Loss probability
Accessibility	3	3	6	0.50	0.50
Accuracy	6	5	11	0.55	0.45
Authority	2	1	3	0.67	0.33
Cohesiveness	7	5	12	0.58	0.42
Completeness	16	49	65	0.25	0.75
Complexity	11	15	26	0.42	0.58
Consistency	8	18	26	0.31	0.69
Informativeness	4	11	15	0.27	0.73
Naturalness	2	9	11	0.18	0.82
Relevance	3	4	7	0.43	0.57
Verifiability	10	28	38	0.26	0.74
Volatility or stability	1	6	7	0.14	0.86
Using IQ standards retroactively	5	4	9	0.56	0.44
IQ improvement work done	14	3	17	0.82	0.18
No IQ improvement work done	26	77	103	0.25	0.75

Table 4: Descriptive statistics of IQ problems found in the Featured Article Removal Candidate votes (out of 120 candidates, 80 articles lost their Featured Article status)

Furthermore, some actual policies and practices were intentionally kept hidden to reduce the cost of overall quality assurance. For instance, to discourage editors from circumventing the system and providing inaccurate metadata, they were not alerted when they uploaded inappropriate images:

Some images are allowed under fair use—however it also contains some “traps”—such as for buildings and living people—that automatically get marked for deletion. The reasoning is that if people get a “you can’t upload that picture” warning they might pick a wrong

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008).

Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

license instead. Deleting images which we know are copyvios is much easier than detecting copyvios that have been labelled as “My work: GFDL licensed.”

When the beliefs and contributions of different agents contradicted each other and could not be reconciled, basically two options were available: discard all conflicting contributions and deliberately avoid the issue, or present all POV. Wikipedia articles often served as good examples of the latter scenario, although one could observe the former kind of compromises as well:

I have revised most of the article, basically writing it anew, although I tried to preserve everything that had not been poisoned by inter-Slavonic recriminations. . . . I’ve deliberately left a lot unsaid here, including historical grammar.

Why has protected status been removed? The people who wanted Ecnomus mentioned in the first paragraph have not consented with the present version and they will no doubt do their stuff again.—User X

I asked for protection to be removed because I wanted to write something about the battle. It’s only one sentence in the introduction, after all, and it’s not wholly objectionable, so I think we can live with the Ecnomus enthusiasts. Eventually they will go away and the reference can be removed.—User Y

Clearly, there could be an economic argument behind this behavior. Representing alternative POVs in highly contested areas without critical analysis of the content and quality of the argument relieves the IQA agent from validation and of some of the negotiation and editing costs. It also gives partial satisfaction or utilities to the disputing parties, motivating them to continue contributing and using Wikipedia. However, it changes a traditional positivist approach toward encyclopedia construction, which assumes that there is always one truth and a certain predictable level of quality, to a constructivist, “grounded” approach, which assumes that there are always multiple truths and quality, and that they change over time. However, along with objective changes caused by changes in the underlying reality and scientific discoveries, updates and modifications can have subjective motivations as well. Information can be changed or reinvented to align with a particular point of view or to achieve a particular outcome. Garfinkel (1967) described how a decision could come before the information when jurors sought information to justify a previously made decision in retrospect.

We found that editors often signaled their IQ beliefs and philosophies through their home pages. This may serve several purposes. Editors may save time by directing opponents to their IQ philosophy statements to clarify their behavior. The statement may also serve as an individual coordination artifact, helping to provide consistency to the individual editor’s IQ assurance actions and behavior—whether the editor supports the inclusion of school articles in Wikipedia, for instance. It can also facilitate searching and identifying like-minded members and forming associations to prevail in various IQ decision-making votes and to adopt desired policies. Finally, the individual IQ philosophy statements may serve as a basis for the policy statements of such associations.

Interestingly, the formal declaration and differentiation of IQ philosophies through establishing associations may help to reduce IQ variance in Wikipedia and make it more predictable. Whether creating member IQ behavior policies and guidelines through the synthesis and integration of member IQ beliefs or simply coordinating member IQ votes, the associations may help to reduce the variance of member IQ behavior. At the community level, the rivalry among associations may also lead to better scrutiny of the existing Wikipedia IQ policies or of the IQ metrics, policies, and behaviors advocated by each group, and may help in finding points of integration and convergence.

Associations may also contribute to the cost sharing of building an IQ assurance infrastructure. Parts of the competing IQ assurance policies and structures can be reused in the enactment of similar mechanisms at the community level. As a result, administrators receive help from members who might not otherwise participate in such activities.

6. Discussion

If one contrasts the work organization models of Wikipedia and conventional databases, one may notice several important differences (see Figure 4):

- First, IQ assessment and the end-user feedback processes in traditional databases are detached from the information creation process, whereas those processes are integrated in Wikipedia. IQ problems, if found, can be fixed immediately within the same IQ assessment process. Indeed, when editors modify each other's contributions, it implicitly or explicitly involves an evaluation of the quality of those contributions.
- In Wikipedia, the end user and editor roles are merged. Any Wikipedia user can instantly become an editor. The end user is a part of the article creation process. In traditional databases, those roles are mostly distinct.
- The product creation and delivery environments for Wikipedia are the same, and there is no time lag between the creation of the product and its delivery to the user. That also means immediate gratification for the editor, although this applies to both good and bad edits.
- Work coordination in Wikipedia is less formal. The decisions regarding which role to play and how work is divided are made by the individuals themselves. The members of the community can self-identify for jobs in

the collaborative work environment based on their interests, knowledge, skills, and available resources (time). For instance, one member can provide deep knowledge and understanding of the subject, whereas others can contribute good editorial skills.

For uncontroversial, “stable” topics, self-selection also ensures that members of editorial groups are substantially well aligned with each other in their interests, backgrounds, and overall understanding of the topics. Hence, self-selection may allow significant savings in terms of selection and coordination costs (Benkler, 2002). It may lead to successful, strongly motivated editorial groups, and consequently to high-quality articles. For controversial topics, on the other hand, self-selection may produce a strongly misaligned editorial group. It can lead to conflicts among the editorial group members and continuous edit wars, and may require the use of formal work coordination and control mechanisms.

The merging of editor and user roles and having article creation processes visible may cause problems when the user accesses a “raw” or vandalized image of the article. The English Wikipedia community is currently considering several options to alleviate this problem. The options include compiling off-line releases of selected reviewed versions of Wikipedia articles and making only the changes made or reviewed by trusted editors visible to the public. It is not clear, however, whether the community will agree to abandon its current edit policy in favor of some of these options.

Currently, instead of preventing errors, Wikipedia’s IQ assurance model focuses more on quick and inexpensive recovery. The assumption is that because there is no ideal or permanent IQ, IQ problems matter as much as the time it takes to fix them. From this perspective, the current egalitarian feedback-based IQ assurance model of Wikipedia is superior to the model used by traditional encyclopedias, where fixing IQ problems may take months, if not years, instead of minutes (Crawford, 2001; Stvilia et al., 2005b). At the same time, it would not be advisable to use the Wikipedia model in areas where the cost of an error can be very high.

Although Wikipedia allows self-selection for different editorial roles, wiki software does include an interdependence mechanism. By allowing the disputing sides to obliterate each other’s contributions easily, a wiki makes the sides interdependent in achieving their goals and, perhaps surprisingly, may encourage more consensus building rather than confrontation.

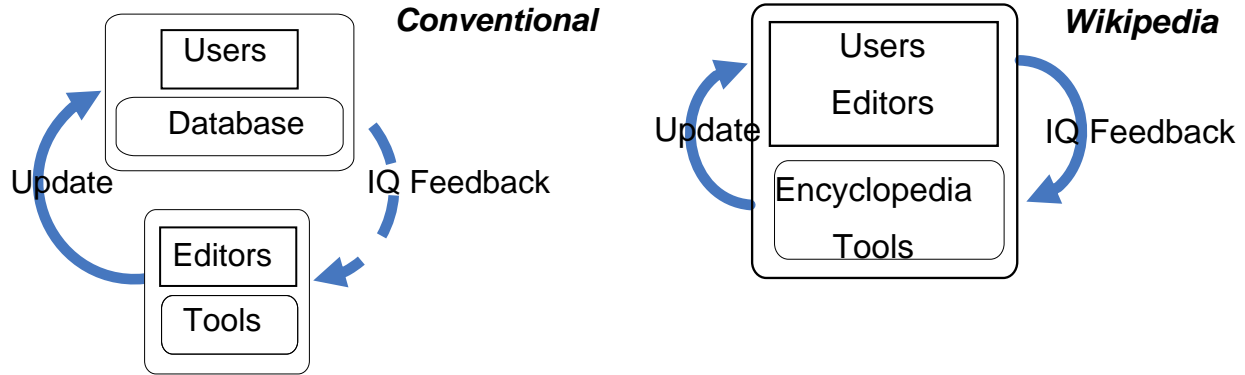


Figure 4: Conventional model vs. Wikipedia IQ assurance model

Furthermore, the community uses additional robust sociotechnical mechanisms for identifying exemplary IQ assurance agents and equipping them with additional system privileges and tools. The privileges make their work significantly less expensive and more effective. At the same time, there is a process for providing feedback to the community regarding an administrator’s performance and for revoking the privileges if they are either not used or misused.

The model of Wikipedia quality assurance evolves with changes in the complexity of the selection task and the use environment. In 2003, Wikipedia had almost no quality assurance guidelines or assessment criteria except for the general “pages that simply will never become encyclopedia articles” and “brilliant prose” statements in the Deletion and Featured Article policies. Within 3 years, the Wikipedia community has managed to develop not only a set of formal quality assessment criteria, but also a whole infrastructure of quality evaluation and promotion mechanisms and guides. New IQ tools and artifacts (bots, WikiProjects, templates, best practice guides) are continually created to promote consistency in the *content*, *structure*, and *presentation* of articles. Indeed, although the article share in Wikipedia pages decreased from 53% in March 2005 to 28% in September 2007, the shares of templates, projects, and, especially, user discussion pages grew significantly (see Table 5).

Page type	March 9, 2005 dump (945,622 pages)		November 30, 2006 dump (5,184,248 pages)		September 8, 2007 dump (8,011,939 pages)	
	Count	Percentage	Count	Percentage	Count	Percentage
Articles	500,623	53	1,774,062	34	2,234,170	28
Talk/discussion	107,569	11	833,766	16	1,544,479	19
User	31,357	3	272,680	5	439,621	5
User_Talk	73,083	8	1,141,370	22	2,226,733	28
Project (Wikipedia)	18,836	2	144,737	3	214,810	3

Project Talk (Wikipedia)	2,870	0.3	14,984	0.3	22,102	0.3
Template	8,477	1	76,161	1.4	99,743	1.2
Template_Talk	1,151	0.1	13,085	0.3	23,667	0.3

Table 5: Page counts by type (does not include redirects)

7. Conclusion

In this article, we presented a case study of the IQ assurance practices used in the Wikipedia project. A qualitative analysis of the quality of Wikipedia articles helped us to understand the ways in which quality is collaboratively established and improved, despite what seems at first glance the seemingly anarchic operation of the project. As a resource for IQ research, a major advantage is that it requires very little effort to obtain compared with other analytic methods such as blind judging. Additionally, we looked at the processes of article creation, and particularly at the article discussion pages as a rich source of qualitative data about what participants in Wikipedia perceive as issues of quality and the processes and trade-offs that operate in activities to improve quality. Results of the study showed that the Wikipedia community takes issues of quality very seriously. Although anyone can participate in editing articles, the results are carefully reviewed and discussed in ways very similar to open source programming projects.

We believe that the study of those evolving debates and processes and of the IQ assurance model as a whole has useful implications for the improvement of quality in other more conventional databases. The classic problems within IQ are trying to determine what quality is, how it might be measured, and what should be done to improve it (Strong et al., 1997; Wang, 1998; Wang & Strong, 1996). Wikipedia offers special insight into these problems. Evolving IQ policies and procedures, assessment criteria, and the surrounding debates and discussions allow us to see how one particular community defines and continually redefines quality, and how this community assesses it in particular cases. Furthermore, the discussion pages reveal how quality issues are discussed and how quality improvements and trade-offs are addressed. What is special about Wikipedia as a resource is that the quality discussions and processes are strongly connected to the data itself and are accessible to the general public. This openness makes it particularly easy for us as researchers to study a particular kind of collaborative work that is highly distributed and that has a particularly substantial focus, not just on error detection, but also on error correction. In contrast, in most conventional databases studied by IQ researchers, these discussions and processes

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008). Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

are divorced from the data, and gaining access to them for research, or even for productive reflection by the practitioners themselves, is very difficult. We believe it will be very interesting and productive to explore how, inspired by the success of Wikipedia, it might be possible to connect quality discussions and processes with the resultant data, and so allow the users of these data various ways to be directly involved in the quality improvement process (Twidale & Marty, 1999, 2000).

Wikipedia and the Wikipedia community would not be possible without a technological innovation, wiki software, which enables very low-cost, collaborative content creation and effective quality control by using a regular Web browser and simple markup language. Using this software, Wikipedians not only create the content of Wikipedia, but also evolve, discuss, and document their work processes, including those related to detecting and improving quality. Hence, not only is the Wikipedia technology influenced by the existing social and cultural structures, but the technology itself also influences the community and makes new forms of social interaction and cooperation possible. This interplay between the technology and community mechanisms creates a new system of IQ assurance that is robust and that promotes continuous IQ improvement.

References

- Bailey, K. (1994). *Methods of social research* (4th ed.). New York: The Free Press.
- Benkler, Y. (2002). Coase's penguin, or, linux and the nature of the firm. *The Yale Law Journal*, 112(3), 369-446.
- Bryant, S., Forte, A., & Bruckman, A. (2005). Becoming Wikipedian: Transformation of participation in a collaborative online encyclopedia. In: K. Schmidt, M. Pendergast, M. Ackerman, G. Mark (Eds.), *Proceedings of GROUP International Conference on Supporting Group Work* (pp. 11–20). New York, NY: ACM Press.
- Cosley, D., Fankowski, D., Kiesler, S., Terveen, L., Riedl, J. (2005). How oversight improves member-maintained communities. In: A. Kellogg, S. Zhai, C. Gale, C. van der Veer (Eds.), In: *Proceedings of the SIGCHI conference on Human factors in Computing Systems* (pp. 11-20). New York, NY: ACM Press.
- Crawford, H. (2001). Encyclopedias. In: R. Bopp & L. C. Smith (Eds.), *Reference and information services: An introduction* (3rd ed., pp. 433–459). Englewood, CO: Libraries Unlimited.
- Crowston, K., & Scozzi, B. (2004). Coordination practices for bug fixing within FLOSS development teams. In: *Proceedings of the First International Workshop on Computer Supported Activity Coordination (CSAC 2004)*orto, Portugal.
- Emigh, W., & Herring, S. (2005). Collaborative authoring on the Web: A genre analysis of online encyclopedias. In: R. Sprague (Ed.), *Proceedings of the 38th Hawaii International Conference on System Sciences* (pp. 99a). Los Alamitos, CA: IEEE Computer Society Press..
- Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008).
Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

- Gasser, L., & Stvilia, B. (2001). *A new framework for information quality* [ISRN UIUCLIS—2001/1+AMAS]. Champaign: University of Illinois at Urbana-Champaign.
- Gasser, L., & Stvilia, B. (2003). Using multi-agent models to understand the quality of organizational information bases over time. In: *Proceedings of the NAACSOS Conference*. Pittsburgh, PA:
- Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438(7070), 900–901.
- Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.
- Grudin, J. (1994). Groupware and social dynamics: Eight challenges for developers. *Communications of the ACM*, 37(1), 92–105.
- Gunning, R. (1952). *Technique of clear writing*. New York: McGraw-Hill.
- Juran, J. (1992). *Juran on quality by design*. New York: The Free Press.
- Lerner, J., & Tirole, J. (2004). *The economics of technology sharing: Open source and beyond* [Working Paper 10956]. Retrieved June 7, 2005, from <http://www.nber.org/papers/w10956>
- Lih, A. (2004). Wikipedia as participatory journalism: Reliable sources? Metrics for evaluating collaborative media as a news resource. In: *Proceedings of 5th International Symposium on Online Journalism*. Austin, TX.
- McArthur, T. (1986). *Worlds of reference: Lexicography, learning and language from the clay tablet to the computer*. Cambridge, UK: Cambridge University Press.
- Orlikowski, W., & Yates, J. (1994). Genre repertoire: The structuring of communicative practices in organizations. *Administrative Science Quarterly*, 39, 541–574.
- Pekkola, S. (2003). Designed for unanticipated use: Common artifacts as design principle for CSCW applications. In: *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work, GROUP 2003* (pp. 359–368). Sanibel, FL.
- Pentzold, C., & Seindenglanz, S. (2006). Foucault@wiki: First steps towards a conceptual framework for the analysis of wiki discourses. In: *Proceedings of the 2006 International Symposium on Wikis* (pp. 59–68).
- Raymond, E. (1998). The cathedral and the bazaar. *First Monday*, 3(3).
- Robinson, M. (1993). Design for unanticipated use . . . In: G. de Michelis, C. Simone, & K. Schmidt (Eds.), *Proceedings of the Third European Conference on Computer Supported Cooperative Work (ECSCW '93)* (pp. 187–202). Kluwer Academic.
- Sack, W., Détienne, F., Burkhardt, J., Barcellini, F., Ducheneaut, N., & Mahendran, D. (2004, November 6–10). *A methodological framework for socio-cognitive analyses of collaborative design of open source software*. Paper presented at the Distributed Collective Practices Workshop in CSCW '04, Chicago, IL.
- Scacchi, W. (2005). Socio-technical interaction networks in free/open source software development processes. In: S. T. Acuña & N. Juristo (Eds.), *Software process modeling*. New York: Springer.
- Schmidt, K., & Bannon, L. (1993). Taking CSCW seriously: Supporting articulation work. *Computer Supported Cooperative Work*, 1(1–2), 7–40.
- Schmidt, K., & Simone, C. (1996). Coordination mechanisms: Towards a conceptual foundation of CSCW systems design. *Computer Supported Cooperative Work*, 5, 155–200.
- Smith, L. C. (1989). “Wholly new forms of encyclopedias”: Electronic knowledge in the form of hypertext. In: *Proceedings of the Forty-Fourth FID Congress* (pp. 245–250). Helsinki, Finland:
- Strauss, A. (1978). *Negotiations: Varieties, contexts, processes, and social order*. San Francisco: Jossey-Bass.
- Strong, D., Lee, Y., & Wang, R. (1997). Data quality in context. *Communications of the ACM*, 40(5), 103–110.

This is a preprint of an article published in JASIST: (Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008).

Information quality work organization in Wikipedia. JASIST, 59(6), 983–1001. <http://dx.doi.org/10.1002/asi.20813>)

Stvilia, B., Twidale, M. B., Gasser, L., & Smith, L. C. (2005a). Information quality discussions in Wikipedia [Technical Report ISRN UIUCLIS—2005/2+CSCW]. Champaign: University of Illinois at Urbana-Champaign.

Stvilia, B., Twidale, M. B., Smith, L. C., Gasser, L. (2005). Assessing information quality of a community-based encyclopedia. In: F. Naumann, M. Gertz, S. Mednick (Eds.), *Proceedings of the International Conference on Information Quality - ICIQ 2005*. (pp. 442-454). Cambridge, MA: MITIQ.

Twidale, M. B., & Marty, P. (2000). Coping with errors: The importance of process data in robust sociotechnical systems. In: *Proceedings of CSCW '00* (pp. 269–278). Philadelphia, PA.

Twidale, M. B., & Marty, P. (1999). *Investigation of data quality and collaboration*. Champaign: University of Illinois at Urbana-Champaign, GSLIS.

Viegas, F., Wattenberg, M., & Dave, K. (2004). Studying cooperation and conflict between authors with history flow visualizations. In: *Proceedings of CHI 2004* (pp. 575–582). Vienna, Austria.

Viegas, F., Wattenberg, M., Kriss, J., & van Ham, F. (2007). Talk before you type: Coordination in Wikipedia. In: *Proceedings of HICSS 2007*. Retrieved February 19, 2007, from http://www.research.ibm.com/visual/papers/wikipedia_coordination_final.pdf

Voss, J. (2005). Measuring wikipedia. In: *Proceedings Tenth International Conference of the International Society for Scientometrics and Informetrics*, at <http://eprints.rclis.org/archive/00003610/01/MeasuringWikipedia2005.pdf>.

Wales, J. (2005). *Interview*. C-SPAN. Retrieved October 20, 2005, from <http://www.q-and-a.org/Transcript/?ProgramID=1042>

Wang, R., & Strong, D. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5–35.

Wang, R. (1998). A product perspective on total data quality management. *Communications of the ACM*, 41(2), 58–65.

Wasserman, S., & Faust, K. (1994). *Social network analysis*. Cambridge, MA: Cambridge University Press.

Weick, K. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage.

Wilkinson, D., & Huberman, B. (2007). Assessing the value of cooperation in Wikipedia. *arXiv.org*. Retrieved March 8, 2007, from <http://uk.arxiv.org/abs/cs.DL/0702140>

Winchester, S. (2003). *The meaning of everything: The story of the Oxford English Dictionary*. Oxford, UK: Oxford University Press.

Yin, R. (1988). *Case study research: Design and methods*. London: Sage.