

# Florida State University Libraries

---

2008

## An activity theoretic model for information quality change

Besiki Stvilia and Les Gasser

This article is an open access copy of the publisher's version of record, which is available at <https://doi.org/10.5210/fm.v13i4.2126>.



First Monday, Volume 13, Number 4 - 7 April 2008

---



## An activity theoretic model for information quality change

by Besiki Stvilia  
and Les Gasser

---

### Abstract

To manage information quality (IQ) effectively, one needs to know how IQ changes over time, what causes it to change, and whether the changes can be predicted. In this paper we analyze the structure of IQ change in *Wikipedia*, an open, collaborative general encyclopedia. We found several patterns in *Wikipedia's* IQ process trajectories and linked them to article types. Drawing on the results of our analysis, we develop a general model of IQ change that can be used for reasoning about IQ dynamics in many different settings, including traditional databases and information repositories.

### Contents

[Introduction](#)  
[Research design and methodology](#)  
[Analysis](#)  
[Discussion](#)  
[Conclusion](#)

---

### Introduction

Information is an increasingly critical resource in our modern lives. The quality of outcomes of individual and institutional processes is often determined by the quality of the information that is used. Because of this link between information quality (IQ) and outcomes of decision making and actions, a theory and tools are needed that would allow effective and efficient management of IQ. A number of frameworks and models have been proposed for measuring IQ (e.g., Strong, *et al.*, 1997). To manage IQ, however, one also needs to know how IQ changes, what causes it to change, and how and when to intervene effectively. In this paper we develop a general model of IQ change based on an analysis of IQ process data in *Wikipedia*. Some statistics on the IQ dynamics of *Wikipedia* articles are also presented.

#### Background and related research

A substantial number of works have analyzed the problem of IQ measurement and have proposed sets of IQ measurement criteria (see Eppler, 2003, for a review). Research focusing specifically on IQ dynamics, however, has been scarce. Researchers have agreed that IQ is contextual. Studies have found repeatedly that moving information from one context to another changes how its quality is viewed and evaluated (e.g., Strong, *et al.*, 1997). Orr (1998) developed a basic control theoretic model of IQ dynamics and suggested a connection between information use and IQ. In particular he argued that information use affects quality and quality affects use in a feedback cycle over time, and that more frequently used information units are more likely to grow in quality. However, this suggests more research questions that need to be answered: what are the IQ interaction parameters, and what are the collective dynamics? Ballou, *et al.* (1998) proposed a decision theoretic model of information workflow process optimization consisting of four variables: timeliness, data quality, cost, and value. Several works have explored the efficacy of mining time series data for detection of failures and intrusions in telecommunication and banking systems (see Milek, *et al.*, 2001;

Pelletier and Dasu, 2005, for sample discussions). However, these studies have focused mainly on macro, aggregate patterns of quality dynamics and have not investigated the underlying micro IQ-related interactions: activities, roles, and strategies. Consequently, they might have overlooked or deemphasized some important variables and relations of the IQ ecology that could be exhibited only at the local levels.

IQ is often defined as the degree of usefulness of information or its “fitness for use” for a particular task or activity system (Juran, 1992; Wang and Strong, 1996). Information activities are complex webs of relationships among actions, roles, norms and conventions, including IQ norms or requirements. Activity theory (Leont’ev, 1978; Nardi, 1996; Vygotskii, 1978) allows one to conceptualize these relationships in a holistic, integrated, and systematic way. One of the main tenets of activity theory is the dialectical notion of tool mediation and evolution, which includes a process of continuous development and learning through individual and collective feedback loops. As such, activity theory can provide a holistic theoretical structure for reasoning about IQ and IQ assurance work.

Successful theories and techniques for quality control have been used in manufacturing (Evans and Lindsay, 2005). Frameworks and methods such as Total Quality Control, Six Sigma, or Statistical Process Control provide powerful philosophical principles for IQ assurance (e.g., continuous quality improvement, the process approach), but it is not clear whether these techniques would be directly applicable to modeling and controlling IQ. The difficulty of applying manufacturing quality control models and techniques to IQ can be found in the peculiar properties of information: its lack of physical properties, the context dependency–nonlinearity of information content (i.e., that information conveyed by the whole is not just the sum of its components), the lack of stability, and the nonrandomness of information errors (Stvilia, 2006).

In an earlier work, we began investigating a dynamic IQ model using an agent-based computational simulation (Gasser and Stvilia, 2003). This simulation modeled the process of a collection of agents differentially interacting with a large information base to accomplish tasks that were driven by the agents’ strategic goals. The four types of agents — user, environment, malicious, and IQ assurance (IQA) agents — both use and change individual information units and the relationships among them to execute tasks that achieve strategic goals. Our simulation suggested a nonlinearity of IQ dynamics as agents selectively improved or degraded information through the use of simple strategies. Using empirical data on IQ evaluation and content change in *Wikipedia* articles, we build on the previous research and develop a holistic model of IQ change, which is illustrated in the following sections.

---

## Research design and methodology

This study used the English *Wikipedia*, a wiki-based, open encyclopedia, to identify the general sources of IQ variance and the patterns of IQ dynamics. *Wikipedia* is a general-purpose encyclopedia. As of September 2007 the English *Wikipedia* contained more than two million articles, more than 10 million objects. It had more than 1,300 active administrators maintaining the collection and more than five million registered user accounts. Since 2005 *Wikipedia* has remained within the top 15 most highly used sites. What is special about *Wikipedia* as an IQ research resource is that *Wikipedia* not only allows anyone to edit its articles, but it also maintains and provides public access to the logs of some of its quality assurance processes (see Stvilia, *et al.*, in press), for a detailed analysis of *Wikipedia*’s information processes). All these make *Wikipedia* an excellent environment for studying and analyzing IQ.

The research method used in this study consisted of a combination of (1) a conceptual modeling of *Wikipedia* information processes; (2) a descriptive statistical analysis and an analysis of time series data of article attribute data; and, (3) a content analysis of quality evaluation discussions.

Activity theory allowed us to develop a conceptual model for reasoning systematically about the general context of IQ in *Wikipedia* — a hierarchy of goal-oriented activities, roles, and the integration points of different sociocultural aspects of the activity system. This activity theoretic model then guided us in data selection by suggesting specific processes which might involve explicit evaluation and decision making on an article’s quality.

The data set for this study comprised edit histories and images of Featured Articles [1] (FA; n=715), Former Featured Articles [2] (FFA; n=375), and a random sample of 1,000 articles

from the 30 November 2006 copy of the *Wikipedia* database. We used these data to generate a time series of monthly data points for the number of article edits, the number of article editors, and article length. Several studies have suggested that there may be a connection between the number of edits and the quality of *Wikipedia* articles (e.g., Stvilia, *et al.*, 2005; Wilkinson and Huberman, 2007). One needs to remember, however, that although the number of edits can often serve as an indirect indicator for quality, for controversial articles a high number of edits can also mean edit wars and vandalism.

In addition, we looked at the logs of Featured Article Review (FAR) and Featured Article Removal Candidate [3] (FARC) process logs for the FFAs (332 vote instances/threads). In particular, we used the results of the descriptive and time series analysis of article attributes (the number of edits, the number of editors, and article length), to guide a more in-depth content analysis of specific instances of quality assurance practices and decision-making processes, and to identify sources of quality variance. Coding was performed by the authors themselves using Atlas.ti software. We started with applying an open coding procedure to the samples. Resultant codes were iteratively clustered to develop a classification scheme (Bailey, 1994). We then used the classification scheme to recode the samples.

Finally, the time series analysis helped to identify some of the trends and patterns in the articles' IQ activities. Graphical data analysis techniques such as run sequence and autocorrelation plots were used to test the data for non-randomness and identify the trends (Chatfield, 1989; Cleveland, 1993). We used SPSS software to do statistical analysis of the samples and generate graphs.



## Analysis

The activity theory framework of the human activity system (see [Figure 1](#)) allowed us to dissect and reason about *Wikipedia's* quality assurance work in a conceptually systematic way. It suggested where and how variance could be introduced in an article's IQ as well as the relationships and points of integration among those variances.

To be more precise, there are three levels of human activity: (1) Activity: activity is the atomic unit, collective in nature and driven by a complex motive of which the individual actors are seldom aware; (2) Actions: activity manifests itself in the form of goal-oriented individual actions in which the subject is consciously aware of what he or she is trying to accomplish; and, (3) Automatic operations: actions in turn rely on automatic, routinized operations, dependent on the conditions at hand. There are continuous two-way transformations between these levels: actions are internalized and become automatic operations through repeated practice, on the other hand, actions may also be expanded into novel collective activities (Engeström, 1990). Note, that the hierarchical definition of activity emphasizes the collective, socially distributed nature of work, which itself implies division of labor.

In addition to the hierarchical structure of activity, activity theory also provides a perspective on historical development and learning. It connects internal (mental) and external (physical) activities and an outside reality through a feedback loop consisting of the processes of knowledge internalization and externalization, and tool mediation. The idea is that human activities are vehicles for both cognitive and social development and learning, and tools become carriers of historically accumulated collective knowledge embedded in their structure and rules of use (Kaptelinin, 1996).

Thus, the activity theoretic framework of analysis can help not only in reasoning systematically about the socio-technical and cognitive aspects and structures of information work, but it can also guide the identification and modeling of the structure of variability present in the work, including IQ variance.

- Culture (Language ...)
- Community (Roles, Norms, Rules)
  - Activities
    - Actions & Goals
      - »Operations
    - Artifacts (Objects or tools)
    - Agents

**Figure 1:** The structure of the human activity system.

An examination of *Wikipedia*'s activity system revealed two major IQ evaluation processes: the process of identifying high-quality or exemplary articles, and the process of identifying low-quality or irrelevant articles and removing them from the collection. Because the logs of deleted articles were not accessible to us, in this study we analyzed only the process of identifying high-quality articles and their characteristics. In particular, we looked at FA selection processes in this work.

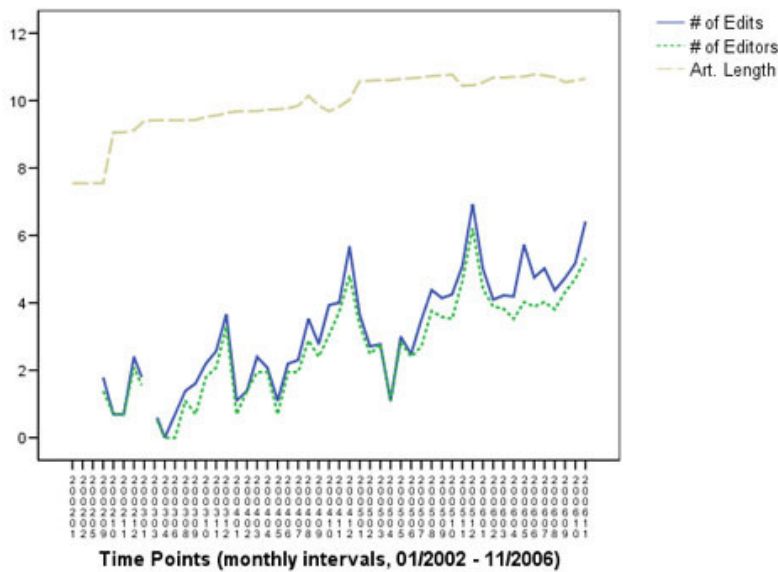
FAs are considered to be *Wikipedia*'s best. According to the FA policy, articles are promoted to FA status by the *Wikipedia* Featured Article Director after the community achieves a consensus that the article meets the FA criteria [4]. Another set of processes (FAR and FARC) is used to demote FAs that no longer meet the FA quality requirements. The FA criteria include both general quality dimensions that are grounded in cultural and social conventions for quality, and characteristics that are specific to encyclopedia article genre and to the *Wikipedia* community. FFAs can be renominated and regain FA status. It is expected, however, that the quality problems identified in the past FARC discussion will be addressed before the article is renominated. *Wikipedia* maintains a list of the FFAs, along with a list of the FFAs that have been renominated and reinstated into the FA collection. The lists, along with the archives of article edits and the related FAR and FARC discussions, contain not only the records of an article's attribute (e.g., number of edits, list of editors) and content change, but also specific instances of an article's IQ evaluation and decision-making. These article attributes can be used as indirect IQ metrics and their changes can be reflections of the changes of the article's IQ parameters or dimensions (e.g., Completeness). Similarly, changes in the community's IQ judgments of the article's quality represent the collective dynamics of its quality dimensions. Connecting and analyzing these two kinds of empirical data can provide valuable information about both a particular IQ assessment model used by the community, and the general structure of IQ change.

The results of the collection level descriptive statistics were unexpected. The editing processes of the FA and FFA articles exhibited similar centrality characteristics and variance, even though the community evaluated the quality of these sets differently, suggesting the presence of additional variance not captured by these measures. Both processes, however, were sharply different from those of the random sample (see [Table 1](#)).

Article type		Mean	Median	Standard deviation
<b>Featured</b> (715 articles, 23,744 time points)	Number of edits	30	7	70
	Number of editors	14	5	29
	Article length	19,481	15,912	16,202
	Number of edits	28	8	64

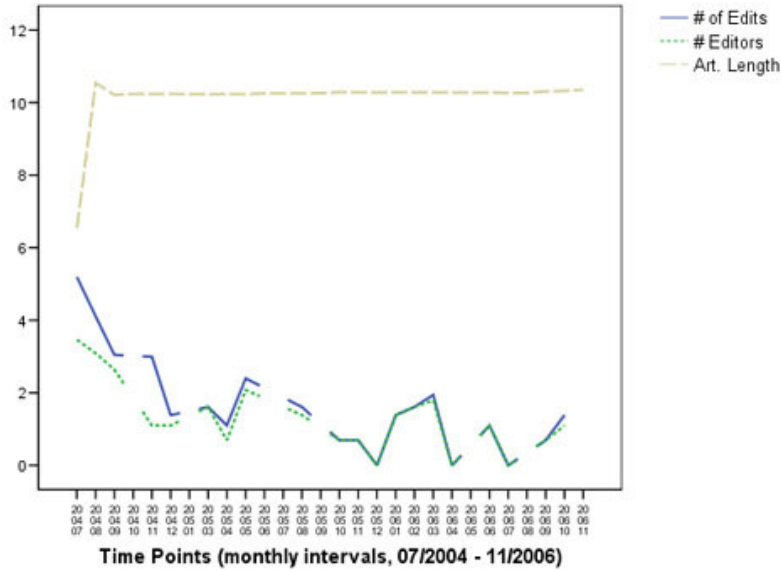
<b>Former Featured</b> (375 articles, 14,260 time points)	Number of editors	15	5	31
	Article length	19,836	16,408	15,500
<b>Random</b> (1,000 articles, 6,012 time points)	Number of edits	3	1	8
	Number of editors	2	1	4
	Article length	3,442	1,764	5,407

The time series data of article attributes were more informative suggesting a connection between the changes in a real world entities and events and the changes in articles' quality. Like many other information objects, Wikipedia articles describe or are about different kinds of entities: *concepts, people, places, events, or things*. [Figure 2](#) shows how the periodicity of a reoccurring event such as a religious holiday may affect the number of edits an article about the holiday may receive. Both the number of edits and the number of editors for the Christmas article exhibit cyclic surges at Christmastime, even though the article length shows less cyclical or seasonal regularity, suggesting that a substantial number of these edits were vandalism or irrelevant additions later reverted by information quality assurance (IQA) agents — *Wikipedia* administrators.



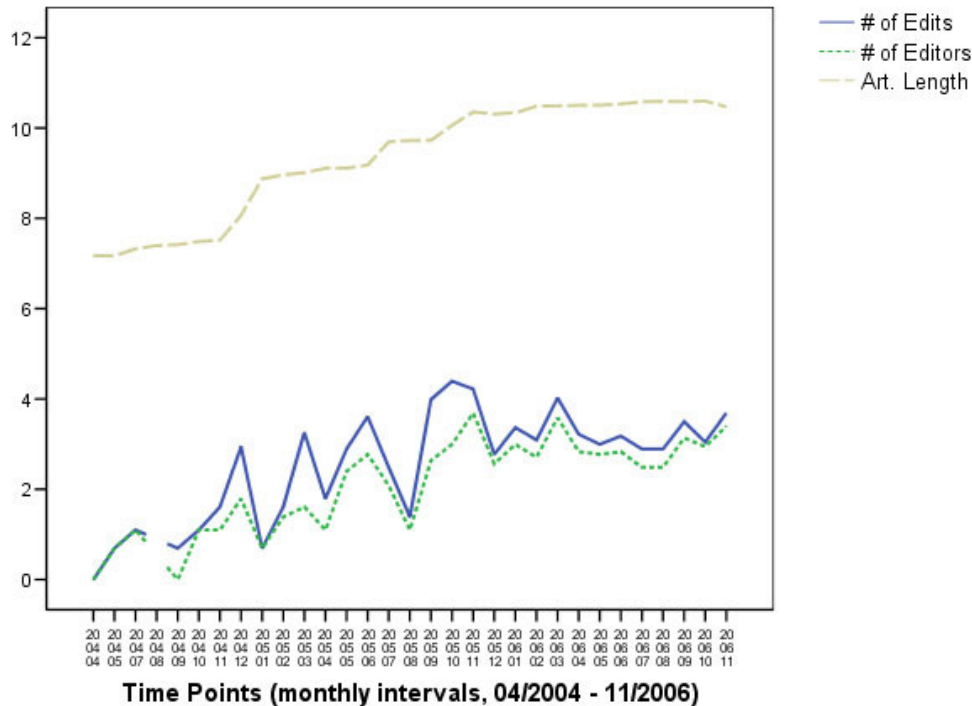
**Figure 2:** Christmas (natural log transformation was used for Y axis).

Article process trajectories about a nonrecurring event or a specific instance of a recurring event, on the other hand, may exhibit a downward trend as more time passes from the time of the event. [Figure 3](#) shows how the monthly rates of edits and editors steadily declined after the event (2004 Democratic National Convention) had occurred.



**Figure 3:** The 2004 Democratic National Convention (natural log transformation was used for Y axis).

The trajectories for articles about persons represent similar trends. Articles about persons who become increasingly famous and influential may exhibit an upward trend for the number of edits and editors (see [Figure 4](#)). Interestingly, one of the picks (October 2005) in the trajectory for an article on Nicolas Sarkozy coincides with the riots in the immigrant communities of Paris, which Sarkozy played a significant role in quelling as Interior Minister. Articles dedicated to persons who ended their career or passed away, on the other hand, may receive less attention from editors unless a person has become a cultural or political symbol whose life is celebrated as a regular event, as in the case of Martin Luther King.



**Figure 4:** Nicolas Sarkozy (natural log transformation was used for Y axis).

Trajectories for articles about concepts, theories, and places did not represent any significant regularities, with the exception of spikes of activity related to FA or FARC processes. In general, religious concepts and theories appeared to attract editors and edits at a higher rate than scholarly concepts.

Interestingly, the article length attribute exhibited less fluctuation in comparison to the number edits and editors attributes suggesting that the cumulative effect of some of these editorial activities might not produce significant changes in the article's content, and the nature of edits might not be homogeneous. This also pointed to the need for more granular, in-depth analysis of the content and structure of edits.

In manufacturing, quality can be improved either by improving the production process (reducing the variance and moving the mean toward the target value through a better process), or by adopting stricter quality control of the ready products (reducing the variance and moving the mean toward the target value through scrap and rework) or increasing the robustness to parameter deviations (Cook, 1997). For digital information products, the line between the production and maintenance processes is generally blurred. This also means that boundaries and attributes of digital information products can be transient. It is easier to modify and recycle a digital information product even after its 'production' process has been completed. *Wikipedia* pushes this to an extreme. At the time of this writing, the concept of article ownership in *Wikipedia* may not apply and anyone can modify articles at any time. The formal distinction between production and maintenance actions disappears, and it may remain in the editors' perception only and be based on how they view their own and each other's edits. As a result, *Wikipedia* articles, even in the FA state, can be treated as "works in progress" and their quality too is expected to be fluid.

In an earlier study we modeled the ecology of a large-scale open information collection by using a multi-agent simulation (Gasser and Stvilia, 2003). All four kinds of information agents modeled in that simulation were found in the *Wikipedia* context as well: (1) editors-agents that contribute or add new content to the article; (2) IQA agents-agents that manage the article and collection quality; (3) malicious agents-agents that purposefully degrade article quality; and, (4) environmental agents-agents that change the IQ of articles through changes



in real-world states. Although mostly degrading IQ, in a few instances changes produced by environmental agents can lead to better alignment of the article's IQ with the real-world state.

Each edit action in *Wikipedia* can be accompanied by the following kinds of actions carried out by IQA agents:

- *Identifying or locating* the contribution;
- *Checking* the validity and quality of contributions;
- *Achieving consensus* contributions through sense-making, discussion, and negotiation; and,
- *Editing* the contribution or the article to better integrate or align the contribution with existing content.

Hence, for each contribution, IQA agents may need to perform one to four actions, on average, per contribution in the quality control activity. Clearly, each of these IQA action types are exhibited through different actions and operations specific to a particular quality problem and activity context (see [Table 2](#)). An extreme case would be a complete reversal or discarding of the contribution. Because the underlying reality described by the article changes over time, the community may need to perform regular *maintenance* or *update* actions to align the article with either the changed underlying entity or the changed general context of article use. Finally, IQA agents improve the process quality by building and maintaining its infrastructure (*i.e.*, developing and maintaining policies and procedures; developing templates, guides, automatic maintenance tools, etc.), including editorial groups, by blocking vandals, resolving disputes, identifying qualified editors, and aligning editors with tasks.

**Table 2: IQ problem types, related causal factors, and IQ assurance actions taken or suggested (FA=Featured Articles; RA=Random Articles).**  
Source: Stvilia, *et al.*, in press

Problem types	Number in FA	Number in RA	Caused by	Action taken or suggested
Accessibility	6	3	<ul style="list-style-type: none"> <li>• Language barrier</li> <li>• Poor organization</li> <li>• Policy restrictions imposed by copyrights, <i>Wikipedia</i> internal policies, and automation scripts</li> </ul>	Reorganize, duplicate, remove, translate, split, join, rearrange
Accuracy	54	53	<ul style="list-style-type: none"> <li>• Typing slips</li> <li>• Low language proficiency</li> <li>• Changes in the real-world states</li> <li>• Wording that excludes alternative points of view (POV)</li> <li>• Garbled by software</li> </ul>	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"> <li>• Lack of supporting sources</li> <li>• Lack of academic scrutiny of the sources</li> <li>• Known bias of the source</li> <li>• Unfounded generalization</li> </ul>	Add, replace, remove, reword, qualify
Cohesiveness	1	1	<ul style="list-style-type: none"> <li>• Loss of focus</li> </ul>	Restrict, move
Completeness	49	20	<ul style="list-style-type: none"> <li>• Existence of multiple perspectives</li> <li>• Unbalanced coverage of different perspectives</li> </ul>	Add, specify, disambiguate, include, expound, balance, qualify, clarify, integrate

			<ul style="list-style-type: none"> <li>Lack of detail</li> <li>Difference between an encyclopedia article genre and the genre from which the text was imported</li> </ul>	
Complexity	7	8	<ul style="list-style-type: none"> <li>Low readability</li> <li>Complex language</li> </ul>	Replace, rewrite, simplify, move, summarize
Consistency	13	12	<ul style="list-style-type: none"> <li>Using different vocabulary for the same concepts within the article or within the collection</li> <li>Using different structures and styles for the same type of articles</li> <li>Nonconformity to the suggested style guides</li> <li>Differences in culture or language semantics</li> <li>Conflicting reports of factual information</li> <li>Contradicting or conflicting with a particular cultural or social norm, convention, or standard</li> </ul>	Reorganize, conform, revert, move, choose the most widely used form, vote
Informativeness	6	4	<ul style="list-style-type: none"> <li>Content redundancy</li> </ul>	Remove, move, revise, cut down
Naturalness	2	1	<ul style="list-style-type: none"> <li>Obscure language; text does not flow well</li> </ul>	Edit, rewrite, improve
Relevance	18	16	<ul style="list-style-type: none"> <li>Adding content that is not relevant or outside the scope of the article</li> </ul>	Revert, move, separate, get rid of, remove
Verifiability	19	12	<ul style="list-style-type: none"> <li>Lack of references to original sources</li> <li>Lack of accessibility of original sources</li> </ul>	Add, remove, cite, revert, provide, confirm
Volatility	2	1	<ul style="list-style-type: none"> <li>Lack of stability caused by edit wars and vandalism</li> </ul>	Avoid, protect

A content analysis of FAR and FARC discussions and votes for FFAs identified three main reasons for changes in their IQ evaluations and their loss of FA quality status (see [Table 3](#)). The analysis showed that 86 percent of FFAs were demoted because of continuously increasing FA quality requirements. The first consistent set of IQ criteria was developed in early 2004 and has been redefined several times since then. The article trajectories reflected these changes by showing surges of editorial activity matching the timing of the criteria changes, and the FAR and FARC review processes triggered by those changes.

It is important to note that the ultimate goal of the FAR and FARC processes is to encourage the existing FA articles to evolve and improve in quality as Wikipedia grows and the supply of FA candidates increases. Most of the time, an article loses FA status if the review process finds

that it does not have the potential to be improved and meet the current quality requirements in a reasonable time frame due to either an inactive or a misaligned editorial group:

This article was featured in November 2004, but currently seems to be in a state of stagnation.

In some cases, editors might disagree with the community consensus about the criteria and simply refuse to make necessary changes:

I am the only editor of this article, and as un-Wiki as it sounds, I wrote it (check the contributions). If I'm not here, it's going to end up out of date (it is already, as it happens). Secondly, I'm very, very annoyed about the requirement for inline citations. When it was made an FA, it wasn't required. Seems like they are now. Well, I know I'm not going to do that.

The second most significant cause of FAs losing their status was modifications of the articles themselves, which degraded their quality instead of improving it. These could be caused by malicious attacks and vandalism as well by unintentional degradation caused by incompetent or irrelevant edits:

If this wasn't a FA, it would get less attention from well-meaning folks trying to "improve" it by adding a link to their favorite fractal gallery.

Finally, changes in the article's underlying entity too could lead to its demotion:

With the introduction of the current S-197 Mustang, and the addition of information and models that was not included in the original FA ... I think the article has been severely compromised.

**Table 3: Causes of FA status removal (332 articles total; more than one reason for losing FA status could be applicable).**

Cause of FA status removal	Number of articles	Percentage
Criteria change	285	86
Article change	47	14
Underlying entity change	3	1

The continuous "work in progress" approach was reflected in the time series data on the number of edits and editors for the articles. Most of the sample articles had a non-zero number of monthly edits. In addition, the time series trajectories exhibited several interesting patterns, which too pointed to a connection between the dynamics of an article's quality and the changes in context or the life cycles of the entities they represented (see [Table 4](#)).

**Table 4: Process patterns for different kinds of articles.**

Trend	Concepts	People	Places	Events	Things
<b>Upward</b>	x	x	x	x	x
<b>Downward</b>	x	x	x	x	x
<b>Cyclic/Seasonal</b>				x	
<b>Flat</b>	x	x	x	x	x



## Discussion

The analysis of *Wikipedia* process logs showed that changes in the IQ of an article were caused by changes in the *article*, changes in its *underlying entity*, and changes in its *activity system’s context* (see [Table 5](#)). Note that the context could change both in *time* and in *space*:

This is en.wikipedia.org (English-speaking world), not usa.wikipedia.org. The article should be re-written to include a WORLD-WIDE view, or it should be de-listed as a featured article.

Furthermore, the process of an information object’s IQ change could be *passive* or *indirect*, caused by changes in the underlying entity and context — culture, sociotechnical structures, and domain knowledge. In general, these changes were not intended to affect the IQ of the object. In the case of *Wikipedia*, for instance, these changes could be a particular editor leaving *Wikipedia* or an article’s editorial group, changes in the FA criteria, or removal or modification of the articles that a given article references or is referred to by. The context could also be changed *actively* to affect the quality of the information object. New sources could purposefully be added or the existing ones modified with the intention of supporting or refuting the information presented by the object (Garfinkel, 1967; Gracy, 2002; Stvilia, *et al.*, 2007). A qualified editor(s) could be invited to help with improving the IQ of an article. There could be *active* or *direct* quality degradation through malicious corruption or removal of the article. Quality degradation actions may not necessarily be malicious, however. We observed in *Wikipedia* how administrators often had to remove edit access to an article (reduce its accessibility) to protect it from greater quality degradation caused by edit wars or frequent vandalism. Alternatively editors might reduce the accuracy of an article by transliterating the phrases written in a script other than Latin to make it more accessible.

Clearly, from the point of IQ assurance, the sources of IQ change (see [Table 5](#)) can be viewed as vulnerabilities the community may need to control. Identifying patterns and trends in the variances of these sources can help in conducting effective pre-emptive intervention and resource allocation. Also, as the time series data of article attributes suggested (see [Figure 4](#)), these vulnerabilities to an article’s quality and the amount of IQ assurance resources spent by the community to address them would increase with an increase in the article’s criticality. Indeed, the analysis of the FARC and FAR data showed that the *Wikipedia* would direct community resources to a particular article in anticipation of an event that could change the quality and/or criticality of the article.

As mentioned earlier, quality in manufacturing can be improved either by reducing the process variance and improving its mean, or by imposing stricter quality control. In *Wikipedia’s* context process improvement can mean formalizing the policies and procedures for article construction, standardizing the style and structure of articles, and better aligning editorial groups with article topics through better communication and selection. The stricter control of the final product would mean more frequent quality review and stricter enforcement of quality criteria. Interestingly, this study found that one of the ways *Wikipedia* improved the quality of its collection was to *continuously increase* the quality requirements for articles to remain in the collection. This not only reduced the collection’s quality variance at the low end, but also increased its mean characteristics without actually changing the production process.

<b>Culture</b>	The culture changes — what was admissible and aligned with the value system of the previous culture may not be admissible or interpreted in the same way in the current culture.
<b>Community</b>	The community makeup as a whole changes — it can become smaller or, larger, more aligned or less aligned, more selective or less selective.
<b>Activities/Events</b>	New activities are introduced that may generate new needs and uses for the information object. Alternatively, some of the existing activities in which the information object was used may become obsolete, making the related information needs obsolete as well. New events may occur that may affect the information object directly (e.g., initiation of a peer reviewing or quality assessment


	process) or indirectly through its underlying entity (e.g., a country has elected a new president).
<b>Agents</b>	Changes occur in editorial groups — existing editors leave or become inactive; new editors arrive who may not be aligned with the group, less qualified, or not interested in contributing faithfully (e.g. trolls, spammers).
<b>Knowledge/Technology/Tools</b>	<p>The current state of knowledge changes — what was considered to be accurate in the past may not be accurate now.</p> <p>New technologies are developed that may change the cost structure for activities, including quality assurance activities — activities that were prohibitively expensive in the past becomes affordable now.</p> <p>Alternatively, a tool or technology may become less effective or efficient with the changed reality, or simply malfunction.</p>



## Conclusion

In this study we analyzed time series data on the edit processes of FAs and FFAs. Although the time series data exhibited different trajectories for different articles, we observed a number of stable patterns in the trajectories. The patterns appeared to follow the life cycles of the underlying entities.

An analysis of FAR and FARC discussions on FFAs showed that IQ could be changed not only actively by editors, malicious agents, or IQA agents editing the article, but also passively by changes in the article's underlying entity or the context of its evaluation and use. The IQ of the majority of FFAs had been re-evaluated as lower, and these FFAs lost their high-quality status after the community decided to increase IQ requirements.

We believe that this study of the patterns of IQ processes and the sources of IQ variance in *Wikipedia* can contribute to a better understanding of IQ dynamics, and that it has useful implications for optimizing IQ assurance in traditional databases. In particular, the activity theoretic model of IQ change and information type specific edit process patterns identified in this study can serve as a reusable knowledge resource for predicting IQ changes and guiding IQ maintenance actions and resource allocation. The model can also inform the design of software architecture and tools for automatic IQ assurance. Future work will include investigating the cost structure of IQ and linking it to IQ decision-making. 

## About the authors

**Besiki Stvilia** is Assistant Professor in the College of Information at Florida State University.  
 Web: <http://mailer.fsu.edu/~bstvilia/>  
 E-mail: bstvilia [at] fsu [dot] edu

**Les Gasser** is Professor in the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign.  
 Web: <http://www.isrl.uiuc.edu/~gasser/>  
 E-mail: gasser [at] uiuc [dot] edu

## Notes

1. [http://en.wikipedia.org/wiki/Wikipedia:Featured\\_articles](http://en.wikipedia.org/wiki/Wikipedia:Featured_articles).

2. [http://en.wikipedia.org/wiki/Wikipedia:Former\\_featured\\_articles](http://en.wikipedia.org/wiki/Wikipedia:Former_featured_articles).

3. [http://en.wikipedia.org/wiki/Wikipedia:Featured\\_article\\_review](http://en.wikipedia.org/wiki/Wikipedia:Featured_article_review).

## References

- K. Bailey, 1994. *Methods of social research*. Fourth edition. New York: Free Press
- D. Ballou, R. Wang, H. Pazer, and G. Tayi, 1998. "Modeling information manufacturing systems to determine information product quality," *Management Science*, volume 44, number 4, pp. 462–484.
- C. Chatfield, 1989. *The analysis of time series: An introduction*. Fourth edition. New York: Chapman and Hall.
- W. Cleveland, 1993. *Visualizing data*. Summit, N.J.: Hobart Press.
- H. Cook, 1997. *Product management: Value, quality, cost, price, profits, and organization*. New York: Chapman and Hall.
- Y. Engeström, 1990. "When is a tool? Multiple meanings of artifacts in human activity," In: Y. Engeström (editor). *Learning, working and imagining: Twelve studies in activity theory*. Helsinki: Orienta-Konsutit Oy, pp. 171–195.
- M. Eppler, 2003. *Managing information quality: Increasing the value of information in knowledge-intensive products and processes*. Berlin: Springer-Verlag.
- J. Evans and W. Lindsay, 2005. *The management and control of quality*. Sixth edition. Mason, Oh.: Thomson/South-Western.
- H. Garfinkel, 1967. *Studies in ethnomethodology*. Englewood Cliffs, N.J.: Prentice-Hall.
- L. Gasser and B. Stvilia, 2003. "Using multi-agent models to understand the quality of organizational information bases over time," *Proceedings of the NAACSOS Conference*, at <http://www.casos.cs.cmu.edu/events/conferences/2003/proceedings.html>, accessed 8 February 2008.
- D. Gracy, 2002. "What you get is not what you see: Forgery and the corruption of recordkeeping systems," In: R. Cox and D. Wallace (editors). *Archives and the public good: Accountability and records in modern society*. Westport, Conn.: Quorum Books, pp. 247–264.
- J. Juran, 1992. *Juran on quality by design*. New York: Free Press.
- A. Leont'ev, 1978. *Activity, consciousness, personality*. Englewood Cliffs, N.J.: Prentice-Hall.
- V. Kaptelinin, 1996. "Activity theory: Implication for human-computer interaction," In: B. Nardi (editor). *Context and consciousness: Activity theory and human-computer interaction*. Cambridge, Mass.: MIT Press, pp. 103–116.
- J. Milek, M. Reigrotzki, H. Bosch, and F. Block, 2001. "Monitoring and data quality control of financial databases from a process control perspective," In E. Pierce and R. Katz-Haas (editors). *Proceedings of the Sixth International Conference on Information Quality*. Cambridge, Mass.: MIT.
- B. Nardi, 1996. "Studying context: A comparison of activity theory, situated action models, and distributed cognition," In: B. Nardi (editor). *Context and consciousness: Activity theory and human-computer interaction*. Cambridge, Mass.: MIT Press, pp. 35–52.
- K. Orr, 1998. "Data quality and systems theory," *Communications of the ACM*, volume 41, number 2, pp. 66–71.
- J. Pelletier and T. Dasu, 2005. "Mining network logs: Information quality challenges," In: F. Naumann, M. Gertz, and S. Mednick (editors). *Proceedings of the International Conference on Information Quality — ICIQ 2005*. Cambridge, Mass.: MIT, pp. 327–339.
- D. Strong, Y. Lee, and R. Wang, 1997. "Data quality in context," *Communications of the ACM*, volume 40, number 5, pp. 103–110.
- B. Stvilia, 2006. "Measuring information quality," Unpublished doctoral thesis, University of Illinois at Urbana-Champaign.

B. Stvilia, M. Twidale, L. Smith, and L. Gasser, in press. "Information quality work organization in Wikipedia," *Journal of the American Society for Information Science and Technology (JASIST)*, and at [http://mailer.fsu.edu/~bstvilia/papers/stvilia\\_wikipedia\\_infoWork\\_p.pdf](http://mailer.fsu.edu/~bstvilia/papers/stvilia_wikipedia_infoWork_p.pdf), accessed 4 March 2008.

B. Stvilia, L. Gasser, M. Twidale, and L. Smith, 2007. "A framework for information quality assessment," *Journal of the American Society for Information Science and Technology*, volume 58, number 12, pp. 1720–1733.

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

L. Vygotskii, 1978. *Mind in society: The development of higher psychological processes*. Cambridge, Mass.: Harvard University Press.

R. Wang and D. Strong, 1996. "Beyond accuracy: What data quality means to data consumers," *Journal of Management Information Systems*, volume 12, number 4, pp. 5–33.

D. Wilkinson and B. Huberman, 2007. "Assessing the value of cooperation in Wikipedia," *First Monday*, volume 12, number 4 (April), at <http://journals.uic.edu/fm/article/view/1763/1643>, accessed 8 February 2008.

---

## Editorial history

Paper received 14 February 2008; accepted 2 March 2008.

---

Copyright © 2008, *First Monday*.

Copyright © 2008, Besiki Stvilia and Les Gasser.

An activity theoretic model for information quality change  
by Besiki Stvilia and Les Gasser  
*First Monday*, Volume 13, Number 4 - 7 April 2008  
<http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/rt/prINTERfriendly/2126/1951>