

# Wiki, Absurd Yet Successful: A Position Paper for CHI 2011 Workshop on Crowdsourcing and Human Computation

Reid Priedhorsky  
Cambridge, Massachusetts, USA  
reid@reidster.net

## THE WIKI MODEL OF COLLABORATION

Over the past decade, a new model of collaborative knowledge synthesis and production has emerged: wikis, which invert the classical publishing model (review happens *after* publication, not before). Further, the wiki philosophy calls for extremely open access (anyone who can read the wiki can also edit it, and these edits are live to all readers immediately upon being saved) as well as transparency (anyone can see who made each change and exactly what was changed).<sup>1</sup>

This model is apparently absurd (“you want me to let anyone who comes to my website edit it, and I don’t even get to check their work before it’s published!?”), yet, with occasional tweaks, it seems to be a huge success. Users do in fact perform the work of creating and synthesizing content as well as the meta-work of reviewing, correcting, and organizing. For example, Wikipedia, the largest and most famous wiki, has amassed over three million encyclopedia articles in English alone [7] and is generally considered to be roughly as accurate as traditional encyclopedias [2]. Wikia, a for-profit wiki company, hosts over 100,000 individual wikis [3]. And mathematical models developed by Cosley [1] suggest that the wiki model develops the same final quality as classical publishing but does so faster.

Yet these and other successes generally focus on textual information (more specifically, collections of article-like blocks of text). There is also significant progress in mapping, with projects like Open Street Map and my own Cyclopath (discussed further below) building large, rich geographic datasets. In other contexts, the success of wikis is less obvious, but the possibilities are tantalizing. Can we build wikis for sketching, animation, or creating structured diagrams? Do wikis work in a Q&A format (current sites like Stack Overflow offer wiki features, but answers are rarely revised by anyone other than the original author, if at all)? Can wikis be used to thoughtfully compare the arguments for and against a contentious legislative proposal (an area being explored by Sen

<sup>1</sup>Much of the ideas and text in this position paper are taken from my thesis [4], research statement, and other previous writings.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Copyright 2011 ACM 978-1-4503-0267-8/11/05...\$10.00.

et al. [6])? Is the wiki model appropriate when a significant portion of the information comes from another source and is not editable (such as annotating privacy policies for websites or analyzing data streams from a collection of sensors)?

I believe the power of the wiki model goes beyond relatively unstructured text. The value of an incrementally improved collective information resource, with low barriers to entry, that affords microcontributions and leverages the power of summarization, is present across many domains. I am eager to explore these ideas, and be challenged, in collaboration with other researchers interested in the power of crowd knowledge.

## BEYOND WIKITEXT

Part of exploring a collaboration model is scoping. Thus, I ask: *what, exactly, is a wiki?* My working definition of the wiki model is as follows. I identify three core and two additional defining properties of “pure” wikis.

- C1. *Post-review.* Changes are live immediately upon being saved, and review happens after publication, not before.
- C2. *Open access.* Any reader, including anonymous readers if such readers have access, can be an editor, and anything can be edited.
- C3. *Transparent changes.* It is easy for everyone to see how the wiki is being changed, and by whom. Specifically, a wiki needs *recent changes lists* summarizing the flow of changes in the wiki or parts of it, *watching* (the ability for a user to be automatically notified when changes occur to a specified subset of the wiki), and *diffing* which lets reviewers see and analyze precisely what changed in a particular revision. This transparency is essential for the monitoring and review tasks which make the wiki model work.
- A4. *Collective ownership.* Each subset of the wiki’s information is owned collectively, not just by the authors of that subset. This is related to property C2: in order for the artifact to have genuine open access, editing must truly be available to anyone, not just the “owners” of a particular territory.
- A5. *Robust consistency.* The artifact’s internal consistency (e.g., the validity of inter-article links in Wikipedia) is difficult to disrupt without explicit intent to do so (small edits are not secretly large edits). For example, the disruptiveness of renaming Wikipedia articles is mitigated

in several ways: the operation is only available to established users, it requires confirming a warning dialog, and users following links to the old name are automatically redirected to the new.

This definition is deliberately narrow, and in fact few, if any, real systems truly meet all five properties. The goal is to clearly define the “pure” wiki model of collaboration as well as dimensions along which wikis vary in the wild; these dimensions can in turn provide useful anchors for researchers.

In my own Ph.D. work at the University of Minnesota, we have applied this model to mapping, creating Cyclopath, a geographic wiki or *geowiki* serving the navigation needs of bicyclists in the Minneapolis-St. Paul metro area of Minnesota. Particular challenges included building a WYSIWYG web map interface with editing (to ensure sufficient ease of use for reasonable open access) and designing *watch regions*, graphical diffing, and other adaptations of the transparency property. Another notable innovation is what we call the *computational geowiki*, where user-contributed wiki knowledge feeds an algorithm – in Cyclopath’s case, route finding, which is actually the most popular use of the system.

The result is a successful production research system. In only a single metro area, Cyclopath has over 2,000 registered users and more anonymous ones, the map has been revised over 12,000 times, and the system has computed over 70,000 routes. The ongoing project has generated one Ph.D. thesis (my own [4]) and six peer-reviewed research papers.

This success story makes it clear that the wiki model can be applied beyond text and beyond simple repositories of information. I believe there is a rich research agenda surrounding this notion; particular directions include:

- We must go beyond Wikipedia. The wiki model is much bigger than any one special case, even one that is highly successful and makes lots of rich data available for study. In particular, we need more production wikis used by real people for real tasks which are highly instrumented by researchers, in order to explore questions like anonymous behavior and interaction at the sub-page level.
- We must explore wikis that are not text-focused. Geowikis are showing promise, and this work should continue. Other possibilities include creating structured diagrams like data models, negotiating multi-person barter deals, and building databases of automated tests for software.
- We must explore wikis concerned with ideas that can’t be practically compared with a ground truth, such as controversial political debates.
- We must explore wiki systems that leverage both machines and people, who are good at different things. Possibilities include using algorithms to increase the value of collected information (e.g., Cyclopath’s route finding) and using mixed initiative to collect the information in the first place, such as integrating data sources like sensors or using wikis to annotate existing non-wiki information (which might itself be changing independently).

- We must measure value from the perspective of information consumers, not producers, even though the latter is easier. The real world offers no points for effort, and one must correctly measure value to effectively obtain more.<sup>2</sup>

To summarize, I am fascinated by the wiki model of collaboration. I am interested in exploring the factors which lead to its success, its limitations, and its applications beyond text information repositories.

## BIOGRAPHY

I am a research staff member at IBM T.J. Watson Research Center in Cambridge, Massachusetts, USA and hold a Ph.D. in computer science from the University of Minnesota. As a researcher focusing on collaborative and social computing, the high-level principle which motivates me is sustainability – I work to empower communities to make better decisions in the pursuit of a sustainable future. I do this by building new tools for creating and communicating knowledge, with a special focus on open content and mass collaboration techniques such as wikis.

In my spare time, I enjoy bicycling, reading, photography, hiking (especially in the deserts of the American West), tinkering and building, and general hacking and programming.

## ACKNOWLEDGEMENTS

Andrea Forte provided thoughtful and challenging feedback. Many of the acknowledgements in [4] also apply here.

## REFERENCES

1. Dan Cosley, Dan Frankowski, Loren Terveen, and John Riedl. Using intelligent task routing and contribution review to help communities build artifacts of lasting value. In *Proc. CHI*, 2006.
2. Jim Giles. Internet encyclopaedias go head to head. *Nature*, 438(7070):900–901, December 2005.
3. Sarah Manley. 100,000 wikis on Wikia, April 2010. [http://community.wikia.com/wiki/User\\_blog:Sarah\\_Manley/100,000\\_wikis\\_on\\_Wikia](http://community.wikia.com/wiki/User_blog:Sarah_Manley/100,000_wikis_on_Wikia).
4. Reid Priedhorsky. *The Value of Geographic Wikis*. PhD thesis, University of Minnesota, August 2010. <http://reidster.net/pubs/thesis.pdf>.
5. Reid Priedhorsky, Jilin Chen, Shyong Tony, Katherine Panciera, Loren Terveen, and John Riedl. Creating, destroying, and restoring value in Wikipedia. In *Proc. GROUP*, 2007.
6. Isaac Sparling. Social media research with Isaac Sparling. <http://www.maclester.edu/news/features/isaacsparling.html>.
7. Wikipedia, January 2011. <http://wikipedia.org/>.

<sup>2</sup>For example, I led a 2007 paper introducing reader-based, rather than author-based, value measurement techniques in Wikipedia [5]. This has generated quite a few citations and sparked a major research initiative in my lab at Minnesota.