

A wiki for the life sciences where authorship matters

Robert Hoffmann

WikiGenes is the first wiki system to combine the collaborative and largely altruistic possibilities of wikis with explicit authorship. In view of the extraordinary success of Wikipedia there remains no doubt about the potential of collaborative publishing, yet its adoption in science has been limited. Here I discuss a dynamic collaborative knowledge base for the life sciences that provides authors with due credit and that can evolve via continual revision and traditional peer review into a rigorous scientific tool.

In WikiGenes (<http://www.wikigenes.org/>), authorship tracking technology enables users to directly identify the source of every word. This was not possible in first generation wikis, although authorship is essential to acknowledge contributors and to appraise the reliability of information. On the basis of clear authorship attribution, users can rate each other, and a self-regulating reputation system can be implemented. This is useful to address quality maintenance and the problem of editing conflicts, which used to depend on slow and theoretically refutable top-down decisions. To facilitate contribution and unambiguous use of scientific language, WikiGenes enables editing of articles in their final layout and citation of scientific terminology and references through integrated database and ontology lookups. All contributions to WikiGenes will be open access.

Dynamic publishing

Wikipedia is global, easy to use, and has a low barrier to access. Many web-based applications have this in common, however. What is outstanding about the wiki model is of course the collaboration of many authors and, moreover, the way in which collaboration manifests itself in the content. A comparison with conventional publications makes this clear.

The medium of conventional publications has changed over thousands of years of human history, but the essential characteristics have not. Conventional publications have a limited number of authors, a precise date of publication, and thus a definite version. In one word, conventional publications are static¹. In science, this means that scholarly discourse must take place in a series of static publications, spread over different journals and media.

Dynamic publications in the wiki model, on the other hand, have theoretically an unlimited number of authors, but more importantly, they have no final version and no definite date of publication. Dynamic articles can evolve with the content and their focus may shift over time. In the context of science, dynamic publications could thus integrate scientific discourse and be potentially always up to date. In a scientific wiki, for instance, there would be no explicit errata, only improved versions of an article.

Continuous integration and harmonization of scientific discourse may not always be desirable and would probably impede scientific innovation and progress. But often it is necessary to synthesize novel insights and theories and to create common reference points. Reviews and textbooks fulfill this function in the conventional system.

The comparison of static and dynamic publishing models shows that they are not mutually exclusive but complementary in the best sense. Dynamic publications, however, are a recent phenomenon, and we are currently exploring whether science can benefit from their integrative potential. This is why so many wikis have been created over the past years in and out of science. By now there are wikis for different scientific communities and on a growing number of topics, ranging from quantum information science to neuropsychiatry. The acceptance of many of these novel approaches in the scientific community has been low, however, and skepticism persists². One reason may simply be that the integrative forces of a bottom-up approach are not useful in all contexts. Most important, first generation wikis (**Box 1**) have not been created specifically for the demands in science, and significant technical innovation is required to unleash the potential of dynamic publishing for scientific discourse.

WikiGenes

WikiGenes is a collaborative knowledge resource for the life sciences, which is based on the general wiki idea but employs specifically developed technology to serve as a rigorous scientific tool. The rationale behind WikiGenes is to provide a platform for the scientific community to collect, communicate and evaluate knowledge about genes, chemicals, diseases and other biomedical concepts in a bottom-up process.

The necessity of this approach originates in the fast-growing body of information in biology and medicine. Over the past decades, a hypercycle of technological advancement, scientific hypothesis and increasing amount of data has led to a shift in perspective toward 'system thinking'^{3,4}. Genome-wide experiments and pathway screenings, for instance, confront researchers with genes or chemical substances they might never have heard of⁵. To sustain scientific progress, it is thus important to provide researchers access to overview information on all biological agents and systems. For this reason, large institutions, such as the National Center for Biotechnology Information (NCBI) and the

Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar Street, Cambridge, Massachusetts 02139, USA. Correspondence should be addressed to R.H. (hoffmann@wikigenes.org).

Published online 27 August 2008; doi:10.1038/ng.f.217

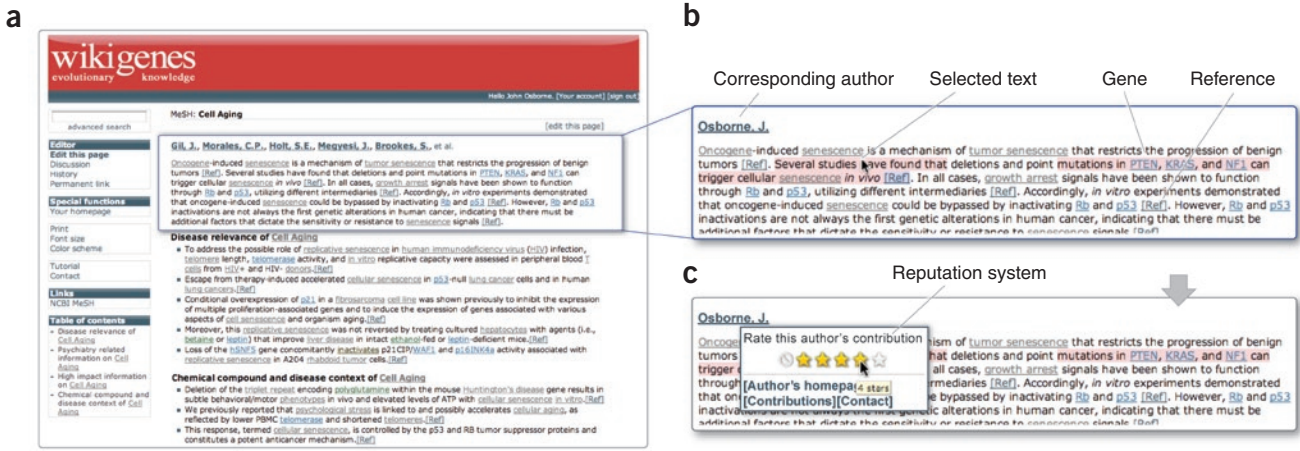


Figure 1 Authorship tracking and reputation system. (a) Front page of a typical article in WikiGenes (<http://www.wikigenes.org/>). Every word in the article is linked unambiguously to its contributing author. (b) Clicking in the text of interest highlights the corresponding author and all of her or his contributions (pink). This way, authors, who invest time and knowledge in their contributions, are given due credit. Readers, on the other hand, can always know the author of any part of a collaborative article in WikiGenes. Hyperlinks on scientific terminology (genes, for example) and references are provided throughout the system. (c) The author name is linked to a context menu, providing access to the author's contributions in other articles and additional information useful to appraise the author. On the basis of this clear authorship attribution, users can rate each other, and it is possible to implement a self-regulating reputation system. This way, readers may acknowledge contributions of particular authors, and dedicated authors could enhance their reputation and assume more responsibility, for instance, in the settlement of editing conflicts.

European Bioinformatics Institute (EBI), invest extensively in expert-curated information resources^{6–9}. These top-down approaches are, however, extremely time consuming and expensive and can ultimately not keep pace—nor scale—with the rapid increase of information^{10,11}. Use of a collaborative community-driven approach, on the other hand, would make it possible to assemble a scientific encyclopedia for biomedicine that would integrate all aspects, be always up to date and evolve with our collective level of knowledge.

Even though there is no doubt as to the usefulness of such a global, community-driven resource in principle^{12–14}, it should not be ignored that the expected effort from contributing scientists would be considerable. This is especially true in a competitive field such as biomedicine, where academic and professional success depends significantly on publication records¹⁵. To make scientific wikis viable, it is therefore essential to exploit all technical possibilities to minimize the necessary efforts for contributors and, more importantly, to provide authors due recognition for their contributions.

Authorship matters

Curiosity and the satisfaction of working together and exchanging ideas are outstanding human characteristics. Another essential element to being a scientist is recognition by others, which translates into employment, grants and, ultimately, the privilege of being a scientist¹⁶. Recognition in science is closely linked to authorship. How many scientists would publish their research and best theories in *Nature* without authorship, anonymously?

In addition to prestige, authorship provides a basis to establish priority of discoveries and theories and to build a reputation among peers. The lack of clear authorship in first-generation wikis (Box 1) is therefore detrimental to the usefulness of dynamic publishing in science. Moreover, in the context of scientific wikis, it is only fair to duly acknowledge authors, who invest time and knowledge in their contributions.

In WikiGenes, authorship tracking technology is used to link every contribution unambiguously to its author, creating the first hybrid of traditional, scientific and collaborative, dynamic publishing (Fig. 1). This technical innovation in WikiGenes also supports the other central

function of authorship as guidance for the reader. Authorship is essential to appraise origin, authority and reliability of information. This is especially important in the wiki model, with its dynamic content and large number of authors. The quality of Wikipedia, for instance, is often so good¹⁷ that it is easy to forget how collaborative articles come about. An author creates a paragraph, another deletes a sentence and inserts a word here and there, and a third author moves a paragraph and adds a new aspect. In brief, the history of a collaborative article can become extremely complex within a few editing cycles.

How could the reader of such an article know who wrote what? In first generation wikis, this information can theoretically be found in the archives and attempts have been made to establish reliability measures¹⁸, but in practice, it is impossible for a user to reconstruct the authorship of specific text passages from hundreds of previous versions.

The uncertainty as to the source of specific texts is therefore an important problem in dynamic publications and decreases the value of articles in their entirety¹⁹. In WikiGenes, on the contrary, new contributions are identified with every editing step and attributed to their authors. Thus readers can always know the corresponding author of any part of a WikiGenes article.

Reputation system and community-driven review

Having a large number of authors is vital to the integration of diverse viewpoints and the efficient assembly of an extensive body of knowledge. Yet, it also harbors the risk of varying quality, tendentious argumentation and, in the worst case, vandalism^{17,19}. In conventional publishing, editorial peer review warrants the adherence to standards, a preselection of information, and a certain level of quality^{20,21}. Scientific wikis should be able to benefit from this time-tested formula.

It is in fact possible to adopt essential elements of editorial peer review in the wiki model. The large number and changeability of dynamic publications and the effort expected from editors and reviewers, however, make conventional peer review difficult to scale. Therefore, it is important that the entire community of authors and readers exerts review and quality control²². In wikis, every author may also act as reviewer, critically questioning or improving disputed information or adding crucial

