Can we trust Wikipedia?

Michael Upshall, Consult MU – "Enhancing Digital Publishing"

michael@consultmu.co.uk

The dominance of Wikipedia as the default reference source is so great that all commercial encyclopaedias have been eclipsed by it. The sheer scale, and the editorial model of Wikipedia that led to its dominance, have rather masked some drawbacks of it. This article suggests some of those drawbacks that any information professional should be aware of.

Scholars and critics have commented on and frequently condemned the Wikipedia editorial model (many of them summarised in *Wikipedia's* own article '[Criticism of Wikipedia'](#)) but paradoxically, the greatest threat to *Wikipedia* as the default reference source for general information is, I believe, the very technology that brought it into being: the Internet, in its latest incarnation as the Semantic Web.

Range

With a print encyclopaedia, every page costs money to print. As a result, even the largest general print encyclopaedias contained relatively few articles: the French *Encyclopédie* had 60,000 articles, and a recent edition of Encyclopaedia *Britannica* 65,000. With over four million articles (Wikipedia 2014), the English language Wikipedia covers more subjects than any earlier encyclopaedia; even so, the number of potential articles appears to be infinite. Although *Wikipedia* guidelines for editors state that only 'notable' topics should merit an entry, there is little agreement on exactly what notable means. In practice, the all-embracing aims of *Wikipedia* mean it is difficult, if not impossible, to resist the inexorable inclusion of additional content. This indicates the impossible challenge that Wikipedia has set itself: in its aim to cover the entire spectrum of knowledge, it cannot set any limits to what is notable. *Wikipedia* is filled, as a result, with articles on topics of marginal interest or value.

The real issue here is quality. Range and quality are of course related. The larger the number of articles, the more difficult it is to curate them, and this seems to be what is happening with *Wikipedia*. *Wikipedia's* own table of Wikipedia article quality ratings reveals that there are over 500,000 entries that have never been assessed by a *Wikipedia* editor. In other words, *Wikipedia* acknowledges it cannot keep up with its own content generation. At the same time, the number of volunteer editors is declining: Wikipedia admitted in 2009 and again in 2012 ([Meyer 2012](#)) that the number of editors and administrators has been declining steadily since 2006.

Quality

Traditional encyclopaedias usually start with a long list of contributors and their academic qualifications – the credentials are often as important as the names. Of course, anyone can edit *Wikipedia*, regardless of ability; the anonymity of contributors makes it impossible to determine who has edited any entry. One of the paradoxes of *Wikipedia* is that registration as a user ensures anonymity more than simply adding or editing content without registration – in the latter case the contributor's identity can be traced. By ensuring anonymity, and not providing sufficient curation, *Wikipedia* is open to allegations of simply representing the views of interested parties; in other words, it may be no more objective than the rest of the Internet.

In the absence of named contributors, *Wikipedia* employs a visible team of editors to review its own content – in public. It is common to see a *Wikipedia* article that has a message attached to it, for example 'This section may require clean-up to meet *Wikipedia's* quality standards'. It has set up a 'Cleanup Taskforce' to deal with inadequate content (Wikipedia 2013). According to its own (not very widely disseminated) quality rating, only around 0.63% of the 4.3 million articles are ranked by Wikipedia itself 'good' or better. An academic study suggests that the quality of articles in Wikipedia correlates with the number of edits they have received ([Wilkinson & Huberman 2007](#)). However, while the authors of this study state 'We also demonstrate a crucial correlation between article quality and the number of edits, which validates *Wikipedia* as a successful collaborative effort', I would argue in contrast that a high level of (voluntary) editorial input cannot be sustained, and an increasing proportion of *Wikipedia* articles will remain without independent editorial intervention. Wikipedia, in other words, is rapidly moving to an agglomeration of articles created and maintained by interested parties promoting a product, person or viewpoint.

Balance and Bias

Perhaps the biggest single problem faced by a traditional encyclopaedia publisher is to ensure balance. Major topics should have the longest articles, and all the articles should follow a similar style. But equally, there should be no consistent political or cultural bias. Such a structure requires substantial editorial capability on the part of the publisher. While one of *Wikipedia's* editorial signposts is the importance of balance, it is well nigh impossible to create balance using thousands of volunteer editors and contributors, all of whom have access to change the content at any time. Even *Wikipedia's* greatest admirers would admit that *Wikipedia* is more an agglomeration of content that will always lack balance, and the consequent lack of authority that this imbalance implies.

A further consequence of *Wikipedia's* emphasis on anonymity for contributors is that without being able to track authorship of content, *Wikipedia* is open to abuse by interested parties writing articles that promote a product or company. Astute readers of Wikipedia cannot have failed to notice the prominent notice at the top of every Wikipedia article during March and April of 2014. Undisclosed paid editing was what Wikipedia is trying to stamp out. The notice states, in no uncertain terms, that if you edit or write articles for Wikipedia, you must by law disclose if you are paid to do so. Will this tactic improve the quality of Wikipedia? It's very unlikely. In fact, it won't make any difference at all. Even though Wikipedia goes on to say the payment might be in kind, or even indirect:

If you work for company Acme, and, as part of your job responsibilities, you edit Wikipedia articles about company Acme, you satisfy the minimum requirement of the Terms of Use if you simply say that you edit on behalf of company Acme on your user page.

This sounds very clear. But the way that Wikipedia has been set up guarantees the anonymity of contributors and editors, so this directive can be ignored with impunity. To prove the point, I signed up to edit Wikipedia. I was told to give a name, and it was suggested to me that using my real name was not necessarily a good idea - fine, I called myself Michael Nobody. After that, I was asked to add an email address: But the email address is optional! I then proceeded to make a couple of small edits to articles (in fact Wikipedia helpfully took me to pages that needed editing). I edited conscientiously, I believe - I changed a capital S for a lower-case S, which is hardly very contentious. I removed the term "passionate" from the sentence: "The **Scottish Youth Parliament**, or SYP, is a democratic, politically independent, inclusive organization set up by a group of passionate youth workers working as a national voice for the youths in Scotland." Who knows, they may well be passionate. But the whole process means that nobody knows who I was. Wikipedia will have a copy of my IP address, but the preference for anonymity for Wikipedia editors means that if I work for Acme, or Johnson & Johnson, or any other company, then nobody need ever be the wiser. I can't see how such a system can ever be changed to improve the quality of entries. Now, if Wikipedia had enough editors to check the articles already posted, it wouldn't be so much of a problem, but as has been pointed out elsewhere, there are more than 500,000 entries in Wikipedia that have never been edited. That's quite a bit of catching up to do.

An earlier version of this article was published in *Culture Unbound,* June 2014.

Michael Upshall has been providing consultancy for publishers on digital content and delivery since 2002. He managed the team that produced the UK's first online encyclopaedia, The Hutchinson Encyclopedia, in 1999. He writes a [blog about reference and encyclopaedia publishing](#).