

What Future for Traditional Encyclopedias in the Age of *Wikipedia*?

Field Report by Michael Upshall

The launch and rapid domination of *Wikipedia* as a reference tool for the Internet was as dramatic as it was unexpected. *Wikipedia* broke so many of the rules of reference publishing, which, even if not formally codified, had been widely accepted for many years: the use of (usually named) authorities as expert contributors, and the presence of moderating editors to ensure balanced structure. All this appeared to have been swept away with *Wikipedia*, and, not least because *Wikipedia* content is given away rather than sold, the competition between *Wikipedia* and most general-purpose encyclopedias was a sad and rather one-sided affair. One by one the existing commercial print general encyclopedias admitted defeat; among the latest is *Brockhaus*, the leading German encyclopedia brand, which ended publication early in 2013.

Of course, 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*' (Wikipedia 2014b), 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. For those unfamiliar with the Semantic Web, it can be defined as 'the exchange of information on the Web via machine-processable data' (Cambridge Semantics 2014), although there are many other, more elaborate and often less precise definitions. What is described as a 'Semantic Web' below is simply the use of automatic tools to pull together content that is more or less related around a common topic. In this paper I examine some of the claimed strengths of *Wikipedia* compared to traditional print encyclopedias, and exam-

ine them in light of Semantic Web developments.

What Advantages do Online Encyclopedias Have?

Range

With a print encyclopedia, every page costs money to print. As a result, even the largest general print encyclopedias contained relatively few articles: the French *Encyclopédie* had 60,000 articles, and *Encyclopaedia Britannica* 65,000. With over four million articles (Wikipedia 2014d), the English language *Wikipedia* covers more subjects than any earlier encyclopedia; even so, the number of potential articles is many more than this. Although *Wikipedia* guidelines for editors state that only 'notable' topics should merit an entry (Wikipedia 2014e), 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 (Wikipedia 2014f) 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.

Topicality

The Achilles' heel of print encyclopedias is always topicality. The work of commissioning content from experts, followed by a critical review, meant that the process of creating and updating an encyclopedia always took several months if not years. The cost of printing means that it is uneconomic to replace an entire volume for the sake of a few updates. When *Wikipedia* was launched, it astonished users because it contained updates from the last few hours. It was as up to date as a newspaper – something unheard of in the slowly moving world of print encyclopedias. Yet *Wikipedia* continues to be updated via a curated model, which means there will always be a delay of several hours from an event occurring and its record in *Wikipedia*. Updates only take place when a user or editor goes into an article and makes a change. In contrast, the Semantic Web model, by publishing dynamically, ensures the most recent updates are immediately available. The Semantic Web will always be more current than a curated model.

Quality

Traditional encyclopedias 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 'good' or better (Wikipedia 2014f). 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 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.

Diderot's *Encyclopédie* did not have signed articles (although the identity of the author has in most cases been identified). Similarly, *Wikipedia* articles are unsigned, and many are composite works by several authors. To compensate for the lack of authority by not having named authors, *Wikipedia* emphasises the importance of citations, and it would seem a valid methodology to try to compel editors to include citations for any claims.

What about quality with the Semantic Web? Intriguingly, the Semantic Web makes no attempt to differentiate content sources; in this sense it is truly democratic. The nature of the Internet means that curated models will become rarer

with time. The Semantic Web is truly democratic, in that no attempt is made or can be made to the user is left to ascertain for him- or herself how reliable the sources are.

Multimedia

Print encyclopedia publishers know that visual material – photos, diagrams and tables – always attracts a disproportionately high attention from readers. Of course, since limitations of space disappear on the Web, an online encyclopedia should outclass any print-based product. Indeed, *Wikipedia* is probably one of the most illustrated encyclopedias available – yet it could be considerably better illustrated. Entries for painters contain at most a handful of their works. *Wikipedia* has a purist approach to content, and tries to keep dictionary definitions in *Wiktionary*, quotations in *Wikiquote*, source content (and many works of art) in *Wikisource*, and so on. For many readers, a valid appreciation of a subject comes via a combination of all of these. In contrast, a Semantic Web mash-up (a dynamically created combination of content from many sources) has no difficulty in including multimedia of many types, such as photos, videos, quotations, definitions, and chemical formulae, as for example in the Learn Chemistry website (<http://www.rsc.org/learn-chemistry>). *Wikipedia* would benefit from displaying its own resources in a mash-up, and by including selected third-party content sites.

Balance and Bias

Perhaps the biggest single problem faced by a traditional encyclopedia 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.

Linking

Traditional publishers have spent many hours attempting to provide cross-references to ensure users are taken as quickly as possible to where the editors have placed an entry: a publisher can place content under 'sea' or 'ocean', but it is impossible to ensure that users always go to the place where the editor chose to put the content.

Many online encyclopaedias, including *Wikipedia*, attempt to solve the problem by converting every example of a word into a hyperlink. Thus, the *Wikipedia* entry for Johann Sebastian Bach states (Wikipedia 2014c) that he was a 'composer, organist, harpsichordist', with organist and harpsichordist as hyperlinks to their respective article. The article for Antonin Dvorak (Wikipedia 2014a) states he was a Czech composer, with 'Czech' being a link. Such a system is easy to implement, but of very limited value to the reader.

Linked data, the expression of relationships in a machine-readable way, is already flourishing in many subject areas, notably life sciences and medicine. One typical use of linked data is to present coverage of a single topic using automatic tools to generate the content. This enables a combination of different media types that *Wikipedia* seems reluctant to attempt. While *Wikipedia* content is available as linked data in the form of DBpedia, this is very different from the creation of a genuine linked reference work.

One idea for reference publishers is to take advantage of the multiplicity of viewpoints and interpretations; for example, *Credo Reference* (<http://corp.credoreference.com/>) do this very well with their topic maps, combining content from several publishers, as well as multimedia. Individual institutions can even create personalised compilations for their users. Of course, some of these treatments may be in disagreement, but the implied acknowledgement that the content is from different providers is, I believe, more sustainable than the *Wikipedia* model.

Wikipedia is not linked data, any more than traditional print encyclopaedias. Every 24 hours, an automatic process is run on *Wikipedia* to extract machine-

readable parts of the content (for example, population figures, dates of birth and death). It is the resulting *DBPedia* that is machine-readable, not *Wikipedia*. The *DBPedia* project, carried out by researchers at the Free University of Berlin and the University of Leipzig, is independent of *Wikipedia*, and only uses a tiny fraction of the total information in *Wikipedia* – that part that can (almost by accident) be converted easily to linked data. It could be argued that the attempts by *DBPedia* to improve the quality of its information, for example *DBPedia Spotlight* (<https://github.com/dbpedia-spotlight/dbpedia-spotlight>), a tool for disambiguation of named entity references, are of more long-term value than all the *Wikipedia* editors.

Topic Page: Nervous system + definition

nervous system, network of specialized tissue that controls actions and reactions of the body and its adjustment to the environment. Virtually all members of the animal kingdom have at least a rudimentary nervous system. Invertebrate animals show varying degrees of complexity in their nervous systems, but it is in the vertebrate animals (phylum Chordata, subphylum Vertebrata) that the system reaches its greatest complexity.

Anatomy and Function
In vertebrates the system has two main divisions, the central and the peripheral nervous systems. The central nervous system consists of the brain and spinal cord. Linked to these are the cranial, spinal, and autonomic nerves, which, with their branches, constitute the peripheral nervous system. The brain might be compared to a computer and its memory banks, the spinal cord to the conducting cable for the computer's input...

Continue Reading

The Columbia Encyclopedia, © Columbia University Press 2013

APA | Chicago | Harvard | MLA Print & Export

Nervous system. (2013). In *The Columbia Encyclopedia*. Retrieved from http://search.credoreference.com/content/entry/columency/nervous_system/0

Images

Nervous system. The nervous system. Nervous System.

Create a Mind Map for Nervous system

Applied Research, Critical Ethnography, Participatory Action Research, Research Hypothesis, Action Research, Critical Action Research, Basic Research, Lifestyle Research, Research.

RELATED TOPICS

- multiple sclerosis
- Animals
- Human brain
- Cerebrum
- Consciousness
- spinal column
- Neurons
- Parkinson's disease
- Pain
- Anatomy

+ See More

Credo Entries | Suggested resources |

Credo Entries

<p>Nervous System <i>The Human Body Book: An Illustrated Guide to Its Structure, Function and Disorders</i> Constantly alive with electricity, the nervous system is the body's prime communication and coordination network. It is so vast and</p>	<p>Nervous System <i>The Brain Book: An Illustrated guide to its structure, function and disorders</i> The nervous system is the body's major communication and control network. Data, in the form of electrical signals, is relayed</p>	<p>nervous system <i>Britannica Concise Encyclopedia</i> A neuron from the visual cortex of a rat Courtesy of Alan Peters System of specialized cells (neurons, or nerve cells) that conduct stimuli from a se</p>
--	---	---

Credo Topic maps: an example of a subject mash-up
Notice that Credo reference does not currently include content from *Wikipedia* or from *DBPedia*, although there is no reason why it should not.

Conclusion: Recommendations for Reference Publishers

In the age of linked data, there remains a vital role for the single-subject curated reference work. Reference publishers can provide these resources with credibility and their limitations of scale make them easier to maintain at a consistent level of editorial integrity that *Wikipedia* cannot achieve. Free but discredited is an improbable business plan.

At the same time, astute publishers will incorporate some (but not all) of *Wikipedia's* editorial model, for example

involving the public in aspects of the content creation and updating, using crowd-sourcing models, for example to suggest updates.

Users will increasingly access reference works via multifaceted websites that take advantage of current technology to combine several different sources, often from different publishers. This linked-data model will increasingly reduce reliance on *Wikipedia* as the default source of reference content via the Internet.

References

- Cambridge Semantics (2014): *Introduction to the Semantic Web*, <http://www.cambridgesemantics.com/semantic-university/introduction-to-the-semantic-we> (accessed 3 March 2014).
- Meyer, Robinson (2012): '3 Charts That Show How Wikipedia Is Running Out of Admins', *The Atlantic*, 16 July 2012: <http://www.theatlantic.com/technology/archive/2012/07/3-charts-that-show-how-wikipedia-is-running-out-of-admins/259829/> (accessed 3 March 2014).
- Wikipedia (2013): *Wikipedia:Cleanup Taskforce*, 4 November 2013: http://en.wikipedia.org/w/index.php?title=Wikipedia:Cleanup_Taskforce&oldid=580212744 (accessed 3 March 2014).
- Wikipedia (2014a): *Antonín Dvořák*, 1 March 2014: http://en.wikipedia.org/w/index.php?title=Anton%C3%ADn_Dvo%C5%99%C3%A1k&oldid=597598579 (accessed 3 March 2014).
- Wikipedia (2014b): *Criticism of Wikipedia*, 2 March 2014: http://en.wikipedia.org/w/index.php?title=Criticism_of_Wikipedia&oldid=597782558 (accessed 3 March 2014).
- Wikipedia (2014c): *Johann Sebastian Bach*, 21 February 2014: http://en.wikipedia.org/w/index.php?title=Johann_Sebastian_Bach&oldid=596528313 (accessed 3 March 2014).
- Wikipedia (2014d): *Wikipedia:About*, 15 February 2014: <http://en.wikipedia.org/w/index.php?title=Wikipedia:About&oldid=595597162> (accessed 3 March 2014).
- Wikipedia (2014e): *Wikipedia:Notability*, 25 February 2014: <http://en.wikipedia.org/w/index.php?title=Wikipedia:Notability&oldid=597077149> (accessed 3 March 2014).
- Wikipedia (2014f): *Wikipedia:Statistics*, 1 March 2014: <http://en.wikipedia.org/w/index.php?title=Wikipedia:Statistics&oldid=597672528> (accessed 3 March 2014).
- Wilkinson, Dennis M. & Huberman, Bernardo A. (2007): *Assessing the Value of Cooperation in Wikipedia*: <http://arxiv.org/abs/cs/0702140> (accessed 3 March 2014).

Michael Upshall was the editor of the UK's first online encyclopedia, *The Hutchinson Encyclopedia*. He was a co-founder of Helicon Publishing, an independent UK print and digital reference publisher, after having edited dictionaries for learners of English (Longman dictionaries), encyclopedias for Dorling Kindersley (the *DK Children's Encyclopedia*) and for Random House. Today he advises publishers on digital publishing strategy and implementation. He blogs about reference publishing at <http://consultmu.co.uk> and <http://withreference.wordpress.com/>. E-mail: michael@consultmu.co.uk