

Many organisations are developing open platforms to create, store and share knowledge. **Alexi Aaltonen** and **Stephan Seiler** analyse editing data by Wikipedia users to show how content creation by individuals generates significant ‘spillover’ benefits, encouraging others to contribute to the collective process of knowledge production.

# Wikipedia: the value of open content production

**F**acebook, YouTube, Twitter and Wikipedia are among the world’s most popular websites – and all of them are based on user-generated content. While some platforms of this kind are primarily used to share individually produced content, others are based on a more direct interaction between users in the production of content.

Wikipedia is the leading example of this type of joint production. The online encyclopaedia contains almost 4.4 million individual articles in the English language version alone, which have been edited by more than 20 million users since its inception in 2001. Wikipedia has largely displaced the former market leader, *Encyclopaedia Britannica*, which is based on a more traditional process of content production (Devereux and Greenstein, 2009).

Wikipedia (and open source

production more generally) constitutes a marked departure from traditional modes of production within organisations. Rather than using a fixed set of procedures to arrive at a pre-specified output goal, open source is characterised by commons-based peer production, a process that is ‘decentralized, collaborative, and non-proprietary; based on sharing resources and outputs among widely distributed, loosely connected individuals’ (Benkler, 2006).

Despite a rising number of products and online platforms relying on this type of production process, we still have relatively little understanding of what drives the growth of content in such environments. Lessons from what makes Wikipedia successful can inform open source projects and ‘wiki’-style platforms in a wide range of public and private sector organisations involved in research, education and innovation.

Without  
‘spillover  
effects’, the  
growth in  
Wikipedia  
editing activity  
between 2002  
and 2010  
would have  
been halved

### Spillovers in content creation

Our study analyses a central question in the context of open content production: does individual content creation ‘spill over’ onto subsequent content creation by other users on the platform?

In contrast with traditional modes of production, it is in the nature of the Wikipedia production process for spillovers between users to occur. Having a large pool of potential editors allows individual contributors to add small pieces of information to an article and rely on subsequent users to develop the content further. How important such effects are quantitatively is the empirical question that our research aims to address.

In contrast to a more traditional editorial process, a Wikipedia user does not need to provide the entire content on a particular topic. Relying on a managerial structure explicitly to organise and coordinate the editing activity is also unnecessary. Instead, a large set of anonymous users interacts in the creation of content. A change in article content might influence other users by providing new information about a topic or by making potential areas for further contributions salient to them, thereby inspiring them to contribute further to the article.

In our research, we estimate the magnitude of these spillover effects and quantify their role in the growth process of Wikipedia content. We are able to do this because of the availability of very detailed information on editing behaviour on Wikipedia. The platform stores the entire history of edits on every article, which allows us to track the evolution of content over time.

We focus our attention on Wikipedia articles that mirror the efforts of more traditional encyclopaedias, namely the incorporation of a given level of knowledge into online content. To this end, we analyse the subset of Wikipedia articles in the ‘Roman Empire’ category, for which knowledge is presumably relatively stable over time.

### Analysing Wikipedia edits

Analysing these data over an eight-year period, we look at how weekly editing activity, measured by the number of

weekly users, is influenced by cumulative past editing activity, measured by article length at the beginning of the respective week. We find a positive effect of article length on editing activity that is statistically significant and economically important.

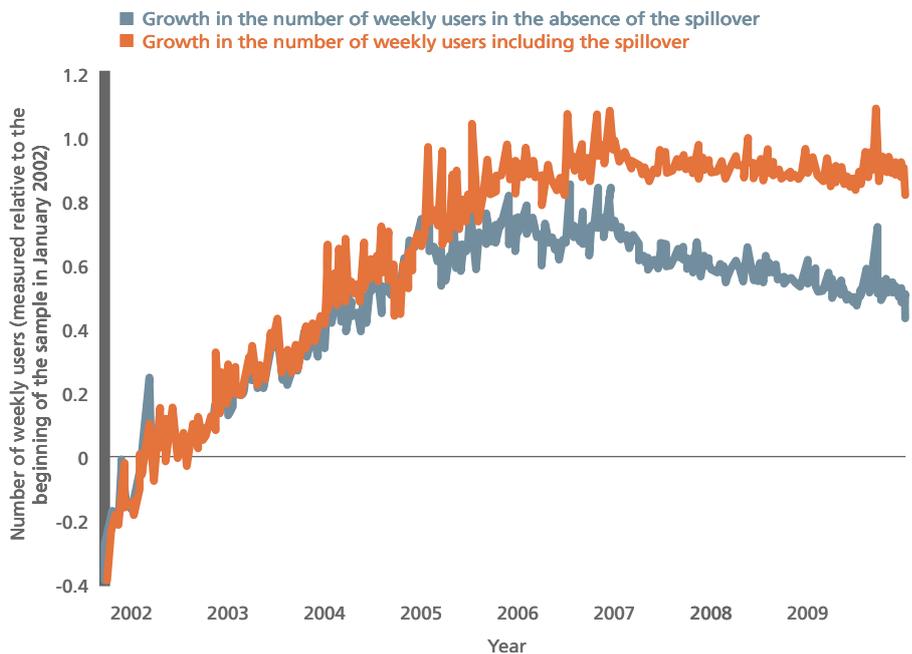
Using the predictions implied by our framework, we quantify growth in editing activity in the absence of the spillover effect to assess its role in the overall growth process. Removing the spillover, we find that the growth in editing activity between 2002 and 2010 would have been halved (see Figure 1, which shows increases in the number of users relative to the first week in the sample in 2002).

More specifically, articles created in 2002 (the only ones that experienced the full growth process) would have had a substantially lower number of weekly users per article in the absence of the spillover. The difference in the growth trends becomes more pronounced over time as articles grow longer and it is strongest at the end of our sample period.

Moreover, article length leads to more editing activity by increasing the number of users editing a particular article. But we find no evidence that the length of edits changes as articles grow. Edits on longer articles are more likely to involve deletion of content and they are more likely to be



Figure 1: Growth in the number of weekly Wikipedia users with and without the ‘spillover’





length and editing activity across articles to the spillover effect whereas in reality it was caused by differences in the popularity or contentiousness of the topic.

Second, Wikipedia as a whole has experienced substantial growth in content over time, which means that in later years, articles will often be longer and edited more heavily. To deal with the first issue, we only use changes in article length for a given article over time, thus eliminating the influence of popularity differences across articles. On the second, we control for the general growth trend across all articles.

### Leveraging spillovers for greater productivity

What are the implications of understanding the growth dynamics and importance of editing spillovers on open platforms beyond Wikipedia? Many firms, including Intel (Intelpedia) and British Telecom (BTpedia), are using internal wiki platforms to create, store and share knowledge within the company. Other public open source projects, such as online dictionaries and a collection of open source teaching material, use the same technological platform as Wikipedia.

There are similar initiatives in the realms of medicine – for example, the ‘Open Source Drug Discovery for Malaria Consortium’ and ‘OpenEMR’, an electronic health records and medical practice management application – and science and engineering – for example, the ‘Science Commons’, which allows the dissemination of scientific work outside academic journals. And there are growing numbers of open source projects that involve the production of physical products: one example is Threadless.com, which relies on a large community of over 500,000 people to design and select T-shirts.

Our findings on the impact of spillovers on Wikipedia suggest the value of all such platforms providing incentives for users to contribute content or to ‘pre-populate’ articles with content so as to trigger further contributions. Since we also find evidence that the magnitude of the spillover effect varies with the total number of users active on the platform, it seems that achieving a larger mass of potential contributors is important for these platforms to benefit from a stronger spillover effect.

This article summarises ‘Quantifying Spillovers in Open Source Content Production: Evidence from Wikipedia’ by Aleksi Aaltonen and Stephan Seiler, CEP Discussion Paper No. 1275 (<http://cep.lse.ac.uk/pubs/download/dp1275.pdf>).

**Aleks Aaltonen** is an assistant professor of information systems at Warwick Business School. **Stephan Seiler** is an assistant professor of marketing at Stanford University’s Graduate School of Business and a research associate in CEP’s productivity and innovation programme.

### Further reading

Yochai Benkler (2006) *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, Yale University Press.

Michelle Devereux and Shane Greenstein (2009) ‘The Crisis at Encyclopaedia Britannica’, Kellogg School of Management Case 5-306-504 (<http://hbr.org/product/crisis-at-encyclopaedia-britannica/an/KEL251-PDF-ENG>).

reverted by subsequent edits – but both effects are small. Finally, we find that the spillover effect triggers content contributions of which

75% can be attributed to new users and 25% to users who previously edited the same article.

So that we can be sure that we are correctly identifying the causal effect of article length, we have to control for two confounding factors. First, inherent differences in the degree of interest in topics will have led to some articles growing longer than others while at the same time attracting higher levels of editing activity. Thus, we might incorrectly attribute the correlation between article

**A larger mass of potential contributors to an online platform will generate bigger spillovers**

**Providing incentives for early users to contribute content will trigger further contributions**