

# Motivation and value of free resources

## Wikipedia and PlanetMath show the way

Aaron E. Klemm

**I**n October of 2000, web-savvy math students lost a critical education tool. MathWorld, an online encyclopedia of mathematics, vanished from the web leaving students, educators, and mathematicians with only a notice that legal problems had caused the shutdown.

MathWorld was an early example of useful web sites for education. Eric Weisstein, the author, originally started it as “Eric’s Treasure Trove of Mathematics.” He spent years collecting and writing entries for what would eventually become a highly regarded reference encyclopedia. As the site became increasingly popular, he struck a deal with CRC Press to publish a print version of his work.

Weisstein had accepted a position at Wolfram Research, and the company offered to help enhance MathWorld and provide hosting for the site. Meanwhile, conflicts with CRC Press began to surface. They wanted to disable portions of MathWorld in order to promote sales of the print version. CRC Press eventually used their contract with Weisstein to claim rights over large portions of his work.

### A replacement emerges

As the MathWorld lawsuit dragged on, several students at Virginia Tech and others from IRC math channels launched PlanetMath, a web site to replace the type of resource Weisstein had created. From the outset they aimed to create a collaborative, community-driven site and chose the GNU Free Documentation License (GFDL) to cover the articles and contributions. The GFDL allows anyone to freely redistribute PlanetMath articles.

“We were all in an essentially defensive mood at the time, after what happened with MathWorld,” Aaron Krowne, a

principal of the project said. “We wanted to ensure that no third party could come along and ‘steal’ the PlanetMath content,” he continues. The GFDL allowed them to do that and guaranteed to contributors that the PlanetMath staff would not unfairly profit from their work. The license allows the authors to retain rights to their own contributions.

Fig. 1: A lawsuit shut down MathWorld



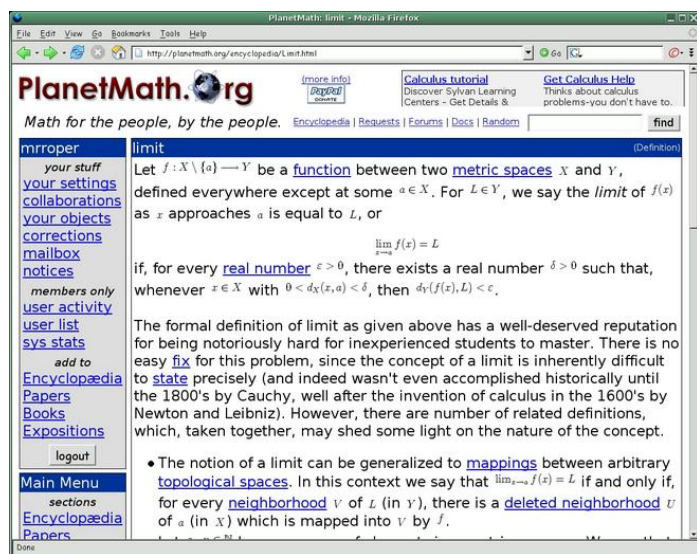
Interestingly, there are plans to create a print version of PlanetMath as well. The GFDL ensures PlanetMath will not encounter the problems MathWorld did.

Krowne was heavily influenced by Yochai Benkler’s paper,

“Coase’s Penguin, or Linux and the Nature of the Firm”. Benkler’s thesis challenges economist Ronald Coase’s belief that production is most efficient in firms and markets.

“I realized that the conditions and attributes that make free software so great are actually just a special case of something Benkler calls ‘Commons-based Peer Production’, or CBPP,” Krowne says. He lists PlanetMath along with GNU/Linux and Wikipedia as examples of what CBPP can achieve.

Fig. 2: PlanetMath is a collaborative mathematical encyclopedia



Benkler writes, “Removing property and contract as the organizing principles of collaboration substantially reduces transaction costs involved in allowing these large clusters of potential contributors to review and select which resources to work on, for which projects, and with which collaborators.”

Since its inception, more than 7000 users have registered on PlanetMath and the encyclopedia section has grown to nearly 4000 entries. Krowne summarizes the benefits of peer-production, “I think a great amount is gained by blurring the line between producer and consumer of content.”

### How much can users produce?

Wikipedia is blurring the lines of production with astounding success. Edited entirely by volunteers, the collaborative online encyclopedia has grown to over one million articles with versions in more than 40 languages.

Founders Jimmy Wales and Larry Sanger originally started Nupedia, a traditional encyclopedia with expert authors and strict review standards to ensure article quality. Nupedia

was innovative only in that it was published on the web for readers at no charge.

After about a year of work, just twenty-four articles were complete and funding was drying up. During that time, Sanger discovered Wiki technology that allows collaborative document editing of web sites. He implemented it to enhance the production of articles prior to submission into Nupedia’s extensive review process. He named the setup Wikipedia and all articles were placed under the GFDL.

As Nupedia fizzled, the developers turned their focus to Wikipedia. The project quickly drew thousands of contributors eager to write about their areas of expertise. Wikipedia now has more entries and is published in more languages than any encyclopedia ever produced.

### Bandwidth is expensive

Wikipedia relies heavily on donations to fund the expensive infrastructure necessary to keep it operating. Three primary costs in creating a project like Wikipedia are development, marketing, and distribution.

Wikipedia and PlanetMath have solved the development and marketing costs by leveraging their unique organizational structure. Contributors are motivated in many ways to help improve the projects: sometimes for recognition, sometimes out of gratitude, sometimes for the challenge, and usually with a commitment to the community they are building. The work gets produced and the proof is visible for all to edit.

Despite the lack of marketing budgets, these projects draw wide interest and are well known on the web. Wikipedia, for example, is now among the top 300 most-visited web sites according to Alexa traffic rankings. Their altruistic and novel nature also generates interest from the press and serves as a no-cost marketing tool.

They wanted to disable portions of  
MathWorld in order to promote sales of  
the print version

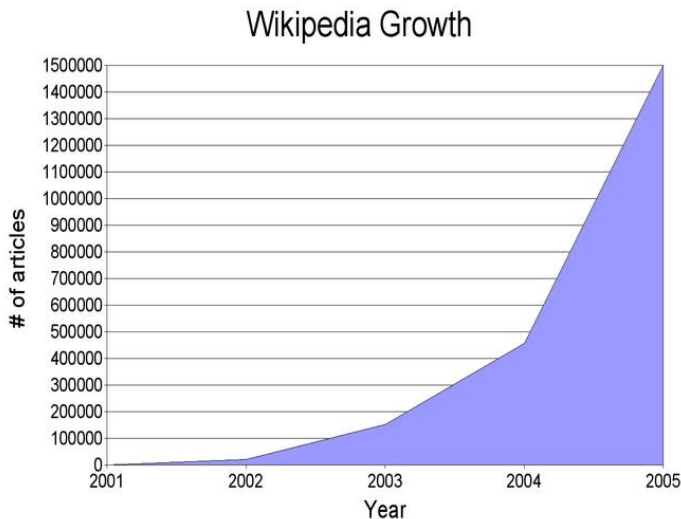
Product distribution can still be problematic for both projects, as servers and bandwidth invariably require time and money to keep serving pages. Wikipedia and PlanetMath continue to face challenges in maintaining a strong infrastructure. It is a consequence anyone could predict.

Developers and advocates of Free Software, Open Source, and peer-produced projects are often questioned about their profit motives. Critics are reflexively suspicious when they

focus on the no-cost aspect. They wonder about quality, support, and longevity.

This author, when explaining to a co-worker his insistence on finding a freely-licensed font, was asked, “How do you expect the font developers to get paid?” A programmer friend suggested this response: “Are you a font developer? Of course not, so don’t worry about it.”

Fig. 3: The number of Wikipedia articles each year since 2001



It was tempting to fire back with that response, but the witty retort would have been a day late, and the co-worker’s genuine concerns about sustainability would have gone unanswered. What the co-worker missed, of course, was that these projects are free of restrictions as well as cost. Relinquishing exclusive, restrictive ownership invites creative minds to extend the ideas implemented in a given project.

“I think a great amount is gained by blurring the line between producer and consumer of content”

Without generating funds directly from the product being produced, free sources (a phrase used for the remainder of this article to mean Free Software, Open Source software, or textual works licensed for freedom of re-use) rely on altruistic contributions, donations, and increasingly more funding from commercial companies that have some interest in further improvement of the project. GNU/Linux has been a significant example of the latter funding model.

Economically, the value opportunities created by Wikipedia are directed outward rather than inward. Third-parties are

given the unique opportunity to profit from works they did not develop. These uses of the Wikipedia database are becoming more visible with interesting results

### I’ll take that, thank you

At first glance it may appear that third-party use is going the way a critic of free sources might expect — leeching. Thefreedictionary.com, by Farlex, Inc., mirrors Wikipedia’s database and places advertisements into the articles as a way to generate revenue. This kind of use is allowed under the GFDL, and Farlex does not technically violate the licensing terms.

The spirit of the license may be under attack in this case, however. Farlex attempts to manipulate search engine results in order to rank their copies of Wikipedia articles higher than those on Wikipedia itself. Farlex relies on this result to increase the chances of site visitors clicking on their ads.

When a user lands on Farlex’s site rather than Wikipedia, several things happen that do a disservice to Wikipedia. First, Farlex does not include an option to edit the page thereby eliminating the possibility for readers to improve the original text. Second, the information Farlex uses is necessarily older than the up-to-the-minute text on Wikipedia. Finally, while GFDL notices are in place on Farlex’s site, the sense of community development is stripped from the text.

Their altruistic and novel nature also generates interest from the press and serves as a no-cost marketing tool

It is a consume-only proposition for site visitors and a “click-our-ads” proposition from Farlex. Wikipedia users are not overly worried about the problem, but they have proposed a variety of solutions. If Wikipedia’s own entries continue to be the most sought-after results, the way linking works on the web may prevent lower-quality links, such as those to Farlex’s site, from jumping ahead in search results. A Wikipedia user concerned about the clone problem recently wrote, “As Wikipedia content proliferates, Google users are going to get more and more annoyed when they do a search and find 15 URLs of cloned material in the top 30 results. As a result, Google will have no choice but to fix this problem eventually.”



As Wikipedia contributors debate that problem, Farlex continues to gain — possibly at the expense of Wikipedia. Is leeching the best free sources can expect from third-party users?

## Free helps free

Programmers are thinking of newer and better ways to integrate web resources into desktop applications, their own web applications, and other software. A software program known as Beagle, from the developers of the freely-licensed Gnome desktop, takes the concept of aggregating disparate resources to a new level. Beagle is not yet officially released with the Gnome desktop, but the code is under very active development.

Beagle tracks a computer user's activities in applications such as email and instant messaging and attempts to derive — from keywords, file types, and other clues — the context of this activity. Then it scans old email, discussions, and documents, which it displays on the desktop for easy and timely retrieval.

For example, while a user drafts an email to a friend about a great new band, Beagle will pull up a list of music files by the band, previous emails that mention the band, and perhaps search results from Amazon about the band's CDs.

In order to make the extension from email archives and the file system across the internet to Amazon and perhaps Wikipedia, Beagle developers must first determine which web resources will allow that kind of use. The Beagle developers have had good luck with Amazon because the company actively promotes web services for external use of their data.

There are plans for Beagle to support encyclopedia lookup features in the future. The developers can be sure licensing problems will not stop them if Beagle starts allowing contextual searches of Wikipedia's vast repository of GFDL articles.

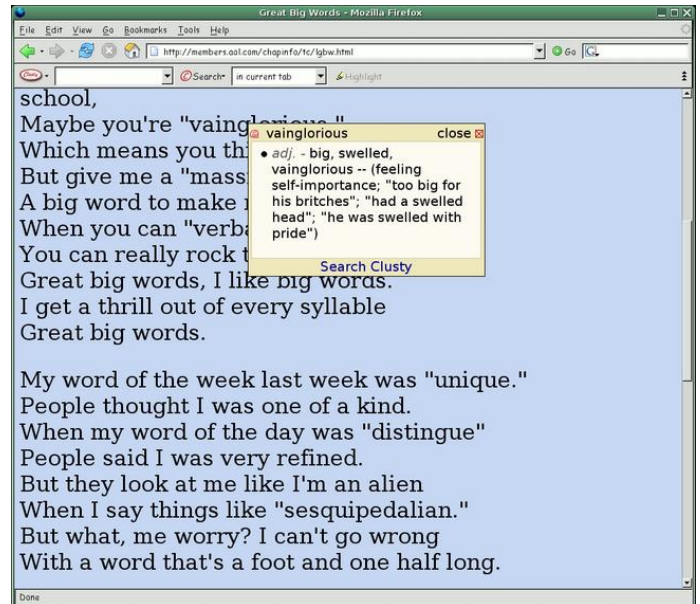
If Beagle has a profound influence on the attractiveness of Gnome (attractive Gnomes?), each third-party currently profiting from the free desktop system will receive a boost. Red Hat and Sun, who have already have successful profit strategies around Gnome, will indirectly benefit from Wikipedia. It will make their products an easier sell.

## A responsible hybrid

The Clusty search engine from Vivisimo, Inc. uses Wikipedia articles directly in their system. Clusty searches

various data on the web and creates “clusters” of links to help organize query results. They mirror a complete copy of Wikipedia for this purpose.

Fig. 4: Clusty's Firefox toolbar enhances page viewing.



“Like the other information sources available on Clusty, Vivisimo's clustering technology helps users find the desired results while discovering unexpected relationships,” Marco Arment, a software engineer at Vivisimo, said.

As an example, Arment explained that a search for “natural language processing” on Clusty will bring up the related articles with links to Wikipedia articles as well as related, clustered results. In one of the clusters is a link to the Association of Computational Linguistics which is something that does not appear in the Wikipedia article.

When a third-party improves upon the work of a free source, that innovation can be re-implemented into the original work. Arment believes third-party users often become contributors to the projects they borrow from and help in other ways. “Third-party use can also highlight new uses for the resources that the creators may not have considered, leading to future expansion and improvement,” Arment said.

Clusty provides a custom toolbar for the Firefox web browser. The toolbar interfaces with Clusty and allows right-clicking on any word in a given web page to access encyclopedia and dictionary “clips” related to the word. A definition or summary of encyclopedia entries pops up in a small window containing contextual links for that word.

Arment thinks free sources provide Vivisimo with unique opportunities to innovate. “In addition to cost savings, free sources give us greater flexibility to use the data in innovative ways. If Wikipedia had a restrictive license, for ex-

ample, licensing conditions and limitations may prevent us from presenting Clusty Clips with the Firefox toolbar,” he said.

“We’re providing additional value to the Wikipedia content by increasing its availability, and we have countless ideas for future expansion of this concept,” Arment said.

## Value, value everywhere

Despite the critics and because of their structure, free sources continue producing quality work. Value opportunities crop up around Wikipedia with no extra effort required by Wikipedia. All players are invited to explore the exploitation of these opportunities.

Because the economic value creation of these projects is not directed inward toward their own sustenance, should we worry that these projects will degrade? If that happens, the value third-parties rely on will degrade along with the project. Responsible third-party players will not be able to ignore this incentive to help sustain free sources.

As each new project proves it can at least exist — a low bar that some critics have seen as too high for the short legs of free sources — third-parties with different goals gain a stake in sustainability. They become part of the community in the

same way contributors and users do.

As Marco Arment puts it, “Every third-party use of a free resource also lends credibility to both the resource itself and the ‘free’ concept in general.”

## Copyright information

© by Aaron E. Klemm

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is available at <http://www.gnu.org/copyleft/fdl.html>

### About the author

Aaron Klemm is an advocate for freedom of software and knowledge resources. He is co-founder of Mathforge.net.