Journal of Physics: Conference Series **331** (2011) 082003 doi:10.1088/1742-6596/331/8/082003

HEP Outreach, Inreach, and Web 2.0

Steven Goldfarb

Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA

Steven.Goldfarb@cern.ch

Abstract. I report on current usage of multimedia and social networking "Web 2.0" tools for Education and Outreach in high-energy physics, and discuss their potential for internal communication within large worldwide collaborations, such as those of the LHC. Following a brief description of the history of Web 2.0 development, I present a survey of the most popular sites and describe their usage in HEP to disseminate information to students and the general public. I then discuss the potential of certain specific tools, such as document and multimedia sharing sites, for boosting the speed and effectiveness of information exchange within the collaborations. I conclude with a brief discussion of the successes and failures of these tools, and make suggestions for improved usage in the future.

1. Introduction

CERN is widely hailed as the "Birthplace of the World-Wide Web" or the "Home of the World-Wide Web". While the former is certainly true and well documented [1], the latter statement is not only wrong, but also highly misleading. Following the historic creation and early implementation of the standards and applications of Tim Berners-Lee and Robert Cailliau, the entire world (as its name implied) quickly became the true residence of the healthy, but rapidly growing, infant web.

Such a childhood is natural for ideas born at CERN. The laboratory is supported by public money, and its products are ultimately returned in the form of discovery, invention, or development, and subsequent documentation. Although these contributions to the body knowledge typically come in the form of publications in academic or technical journals, public knowledge of the web first came in 1991 less formally, as postings to a pair of Internet discussion groups [2]. These were followed up with a presentation at CHEP 1992 in Annecy [3]. At the time, only 26 web servers had been deployed, primarily in Europe. The following year, however, more than 200 servers were installed, followed by an explosion of development that came on the heels of the Mosaic browser release in 1993 [4]. Anyone with access to a web server, browser, and a text editor could create web pages. Development was only limited by the world's imagination and, so far, there appears to be no end to that.

2. Web 2.0

2.1. What's In A Name?

Coinage of the term Web 2.0 can be attributed to an article written by Darcy DiNucci in 1999 [5], referring to the proliferation of the World-Wide Web into processors located in household devices. The original article is protected, missing or impossible to find (confirmed by others who have gone down this path [6]), but many agree on this as the first reference. That is not, however, today's

Published under licence by IOP Publishing Ltd

common usage, nor is it the topic of this note. Rather, I refer here to the iterative usage of web-based tools to develop content for the web; that is, the Web 2.0 that was the topic of discussion in 2004 during the first Web 2.0 Conference, hosted by John Battelle and Tim O'Reilly in San Francisco [7].

By shifting development and maintenance of web content from traditional desktop applications and editors to web-based tools, the number of potential contributors explodes. The unique characteristics attributed to Web 2.0 are denoted in a short, but informative article by Prashant Sharma [8]. To briefly summarize, Web 2.0 has the effect of enabling many more users to contribute to content, since the publishing and editing applications reside on the web server, and are thus available to anyone with a browser. Additional contributors provide for faster development, more access to ideas and data, automatic (and sometimes unsolicited) peer review, the possibility for large-scale collaboration, and the inevitability of increased competition, yielding more and better products. The server-based tools also have an advantage over traditional desktop tools, in that they also benefit from the availability of large-scale computing and storage, as well increased usage and feedback.

2.2. A Strong Back Bone for Fast Growth

From this fertile environment, a rich variety of sites appeared on the scene. The successful ones were those that provided a relatively simple interface, backed by sufficient computing and storage to accommodate the world as a development team. The support structures include wikis, for the simple posting and editing of text and images; blogs, allowing users to collaborate and exchange ideas; multimedia posting sites, for the storage and sharing of image and video files; and social networking sites, allowing individuals or groups to aggregate and share information about themselves or whatever topic interests them.

Behind the scenes was the introduction of tools that facilitate development, perhaps the most important of which being the CMS or Content Management System. These systems enable the management of content by allowing users to enter, manipulate and cross-reference data and metadata. They came about naturally as developers converged on the usage of templates (often XSLT), common scripting platforms (usually PHP), and the storage of data and metadata in relational databases (typically MySQL). Usage of public domain or Open Source systems allowed for the rapid and collaborative development of these systems and a large number of competing systems exist today [9]. The world, as a developer, is tough to beat.

3. Web 2.0 Usage for HEP Outreach

3.1. The Advantages of Web 2.0

Given the competition, it is not surprising to see that developers at CERN did not remain at the forefront of web development for very long. So, when Outreach and Education programs began looking for tools to attract the attention of the public – most recently to promote the research of the LHC – they turned to Web 2.0. Not only did these tools provide improved functionality with a greatly reduced effort for maintenance, but also they had become immensely popular.

3.2. HEP in the Limelight

And that popularity can not be ignored. For whatever reason – the scale of the projects, their cost, a newfound popularity of science, increased efforts by the outreach coordinators and press offices, or perhaps the unpredictable success of that "vague but exciting" application called the web – the media and public has taken a great interest in the LHC. Hundreds of thousands watched the live webcast from CERN of the initial circulation of protons in the LHC in October 2008. Millions more read about it in journals and newspapers, listened to the news on radio, watched it on TV and, of course, followed everything on the web. The LHC First Physics event in March 2010, during which high-energy collisions were produced for the first time, drew a worldwide live webcast audience of 2,500,000 and hit major publications (online and print) around the globe [10].

3.3. Staying in the Limelight

Taking advantage of this rising tide of interest is vital to the HEP community. First and foremost, it is the responsibility of academia not only to perform research, but also to inform the world of the results and implications of our studies. Scientific journals are sufficient for keeping our own community informed, but other measures are required for a public that might not be sufficiently trained in the mathematical and scientific jargon to grasp the significance of the progress being made.

Furthermore, if efforts are not made to maintain public interest in our work, support will dwindle, especially during tight economic times. It is a never-ending challenge to remind the public of the long-term rewards provided by fundamental research. While none of us doubt the importance of the message, the questions remain: How do we take advantage of this newfound visibility? What tools out there will help us to best show off our work and to educate the public?

3.4. The Major Players

Over the past few years, many public Web 2.0 sites have come and gone. Those that target human interaction, such as sharing of documents and multimedia material; communication of information or opinions among users; or aggregation of material into portals or profiles, are referred to as Social Networking sites. Of these, the major players currently in extensive use for HEP outreach are Facebook [11], Twitter [12], MySpace [13], Flickr [14], YouTube [15], Vimeo [16], and Wikipedia [17]. In addition, a variety of portals act as hosts to Web 2.0 tools, such as blogs or discussion groups.

Figure 1. CERN Facebook site (October 2010).

Two of the most important features shared by these sites are their "findability" (viewers now naturally turn to those sites to find information) and their "searchability" (a search for a specific laboratory or experiment, using any popular search engine, will find its own public page, followed directly by relevant pages hosted on these popular social networking sites). In addition, those sites provide powerful search mechanisms, allowing one to easily find specific material related to the host laboratory or experiment.

The most popular social networking site is Facebook. More than half a billion people use it. Central to each host site is a "wall" containing a continual blog of messages written by the host or by "friends" of the host. Figure 1 presents a screen capture of CERN's Facebook wall. Messages can contain text, links, images, or video, and the viewer can select authors to follow. Facebook and several other sites have standardized API's allowing for the cross-linking of posts. A message posted on Facebook, can also be made to appear on Twitter or MySpace, if the host chooses to do so. Furthermore, simply

sending a link results in the inclusion of images and text from the linked site. In addition to the wall, there is a "profile" providing a description of the host, and a library for sharing multimedia material. One can choose which material is accessible by the public and which is only for "friends".

For HEP content providers, the most valuable aspects of Facebook are the profile, as a highly visible and easily located portal to descriptions of the laboratory or experiment, and the pool of "friends", giving one access to a large audience for announcements of interesting events, projects, or new material. In fact, the audience is even larger than the number of friends, in that friends of friends can be reached via the walls of their friends. Every time a viewer "likes" or "shares" a posting (by clicking on an associated button), that viewer's friends and their friends, etc. will get the message. Given a large enough following, a popular posting can go "viral". That is, it proliferates quickly through the web to a very large audience.

Twitter is a social networking site that acts as a kind of soapbox for very short speeches. Hosts write postings, called "tweets", of 140 characters or less, and which are seen by any user that chooses to "follow" them. Unlike Facebook, Twitter does not require reciprocity for followers, so anyone can choose to follow your tweets, regardless of whether you choose to follow theirs. This automatically increases the potential audience. In addition, any follower can choose to "re-tweet" a host she is following, dramatically enlarging the potential audience. Given the "six degrees of separation" rule [18], one could argue that six re-tweets would suffice to cover the planet!

MySpace is a general integration site, similar to Facebook, but primarily targeted for musicians. HEP entities promoting musical projects, such as the ATLAS Resonance CD [19], have exploited the site in order to reach a typically younger audience. MySpace also maintains profiles and the ability to broadcast messages to lists of friends, called "fans".

Flickr, YouTube, and Vimeo are multimedia-hosting sites. They allow hosts to publish and share images or video. Flickr is primarily used for images, with a simple web interface for uploading of data and metadata, as well as integrated plug-ins for commonly used applications. YouTube and Vimeo are the most commonly used sites for video content. Hosts set up channels of video content and can easily send announcements of new material to their channel "subscribers". HEP communicators are exploiting these sites more and more, as multimedia material has gained an increasingly important role in in outreach and education projects.

Wikipedia is an information-sharing site. It is essentially an on-line encyclopaedia written by, well, the world. There are groups of qualified reviewers and editors behind the scenes, ensuring some degree of data accuracy and quality, and articles are required to conform to certain rules that support legitimacy (proper references to support material, for example). Many HEP laboratories, experiments and projects are now listed there, entered either by official representatives, collaboration members, or even unknown volunteers from the public. Although (Because?) the whole world is free to edit and to add to content, the articles tend to remain current and remarkably accurate.

3.5. What's Hot?

Table 1 presents a listing of some of the most popular HEP Outreach Web 2.0 sites, including a few of the author's favourites. Certainly the largest measurable audience would be the 240,000 followers of the CERN Twitter site. Recall that this does not include re-tweets, which cannot be accounted for and are potentially quite prolific.

The majority of HEP Outreach webmasters still rely most on traditional web sites to serve as the primary portals for their laboratory or experiment. These sites, however, nearly all integrate a variety of Web 2.0 tools, such as blogs, and usually include direct links to complementary social networking sites. Sites such as Facebook also provide buttons for the standard sites that a viewer can click on to indicate they "like" that particular web page or to "share" that page, thus advertising a home page to their friends, using the Facebook network, and increasing the audience. Other Web 2.0 tools, such as wikis, are simply used because they facilitate the development and maintenance of the content. Which is one of the reasons why HEP internal communication is also interested.

Table 1. Examples of popular Web 2.0 sites with HEP content.

Site	Popularity
CERN Facebook Site	30,000 "like" it
CERN Twitter Site	240,000 followers
Public Flickr Images about Physics	1,800 members (4400 images)
CERNTV YouTube Channel	13,500 subscribers
Best Of Science YouTube Channel	50,000 subscribers
Wikipedia LHC Entry	500 revisions (86 references)

4. Current and Potential Usage of Web 2.0 for HEP Internal Communication

4.1. The Challenge

Internal communication for today's large, international HEP collaborations faces many important challenges, all of which play a role in the effective running of the experiments. The largest of the LHC collaborations, ATLAS and CMS each comprise more than 3000 members, from a list of countries that truly spans the globe. Nearly all of the physical components of the detectors were constructed at remote institutes and assembled at CERN, so clear reliable communication has been essential from the start, to ensure that the coordination of their design, construction, and maintenance. Equally important, has been the development of the software, distributed computing infrastructure, and physics analysis strategies, none of which could rely only on CERN-based teams.

To ensure the success of the construction and running of the experiments, as well as to guarantee equal participation for all parties to the key decision-making processes that drive the experiments, a large variety of collaborative tools have been developed and installed at CERN and at the remote institutes [20]. These include audio and video conferencing, lecture archiving, webcasting, web-based application sharing, and – yes – Web 2.0 tools and social networking sites.

4.2. Current Tools

Web 2.0 tools currently receiving the most usage for HEP internal communication include wiki pages and content sharing tools, such as Microsoft Sharepoint. These systems allow remote contribution and editing of content, using web browsers, and support user authentication. Both also have important shortcomings. Wiki pages, typically implemented using Twiki [21] are easy to create, but are also easy to forget. There are no automatic systems to ensure maintenance, so stale content must be found and removed by hand. The wiki interfaces for the management of files or images tend to be primitive and restrictive, and content is not as findable or as searchable as on the social networking sites described above. Sharepoint sites offer more sophisticated tools for the management of content, but development and maintenance using the web interface is slow and clumsy. Sites can be designed with proprietary software tools, but that work model does not fit well to the heterogeneous community of HEP.

4.3. Current Development

Efforts are underway at CERN to migrate web development to the Drupal [22] content management system. This Open Source system has benefited from its popularity to become one of the most mature tools available, and it comes with a significant number of user-developed modules. Moving to such a system will not only improve the functionality of the CERN web pages, but will also help to interlink content between internal and public pages, as well as across various organizational units in the laboratory. Moreover, Drupal is designed for web-based development, ideal for collaborations working on a variety of computing platforms and located all over the globe.

Deployment of Drupal should address many of the web content use cases prevalent to HEP laboratories and experiments. One can envision a scenario, for example, in which a key LHC event display image enters the system as a protected component of one of the experiment's data analysis chains, then appears as part of a physics approval package, a physics publication, a seminar, and

eventually a key component of a press release, public web pages, and social networking sites; all of this, following one single data entry. Control of authorization using CERN authentication systems are being integrated into the implementation with such scenarios in mind. It is my opinion that usage of the system will only be limited by our imagination, human resources, and our ability to agree on protocol for the various use cases.

5. Up Next: Web 3.0

Yes, of course, the phrase Web 3.0 exists. It is typically used in reference to what has been referred to as the "Semantic Web". That is, the usage of well-defined semantic tagging of content, making it possible for applications to "understand" the relevance of the content, and thus to aggregate material pertaining to a specific task. Of course, as with any database, one still needs to supply data and intelligent metadata. That is, although we will not find the Higgs with Web 3.0, writing the discovery talk might become that much easier. Perhaps that is a better topic for CHEP 2012.

References

[1] CERN documentation on info.cern.ch, the website of the world's first web server.
[2] Berners-Lee, T. *Qualifiers on Hypertext links* and *WorldWideWeb: Summary* were posted to the open internet discussion groups alt.hypertext and comp.archives, 6 Aug 1991.
[3] Cailliau, R, Berners-Lee, T, WorldWide Web; Invited talk at conference: Computing in High Energy Physics 92, Annecy, France, 2327, Sep 1992.
[4] Andreessen, M. NCSA Mosaic Technical Summary, National Center for Supercomputing Applications, version 2.1, 8 May 1993.
[5] DiNucci, D, "Fragmented Future", *Print* **53** (4): 32, 1999 (password-protected, unfortunately).
[6] Ruiz, C. "Who Coined Web 2.0?: Darcy DiNucci", Cole20, 12 Feb 2008.
[7] O'Reilly, T, and Battelle, J, "Web 2.0 Conference Opening Welcome: State of the Internet Industry" San Francisco, California USA, 5 Oct 2004.
[8] Sharma, P., *Core Characteristics of Web 2.0 Services*, TechPluto, 28 Nov 2008.
[9] CMS Matrix, a comparison of both Proprietary and Open Source content management systems curently on the market.
[10] Taylor L, Barney D, Goldfarb S, *ATLAS, CMS and New Challenges for Public Communication*, these proceedings.
[11] Facebook is a social networking site, integrating messaging and mulitmedia management with profile hosting for persons or groups.
[12] Twitter is a social networking site, allowing participants to send messages to their "followers".
[13] MySpace is a social networking site similar to Facebook, but generally focussing on the hosting of music and video, providing profiles of musicians and bands.
[14] Flickr is a social networking site set up for the hosting and sharing of images and video.
[15] YouTube is a social networking site set up for the hosting and sharing of video.
[16] Vimeo is a social networking site set up for the hosting and sharing of video.
[17] Wikipedia is designed for the collaborative development and editing of a global encyclopaedia.
[18] Six Degrees of Separation, wikipedia entry, 9 Jan 2011.
[19] ATLAS Resonance is a musical album released in 2010, comprising songs performed by members of the collaboration, and used to promote the physics of ATLAS to new audiences.
[20] Goldfarb S, *Collaborative Tools and the LHC: Some Success, Some Plans*, *International Conference on Computing in High Energy and Nuclear Physics*, Mar 2009, Prague.
[21] Twiki is an open source enterprise wiki implementation based on templates.
[22] Drupal is an Open Source Content Management System.