



HAL
open science

Web 2.0

Aurélien Bénel

► **To cite this version:**

| Aurélien Bénel. Web 2.0. SAGE Encyclopedia of the Internet, 2018. hal-02363529

HAL Id: hal-02363529

<https://utt.hal.science/hal-02363529>

Submitted on 14 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

WEB 2.0

In 2001, the bursting of the dot-com bubble led some analysts to argue that the World Wide Web was overhyped. A few years later, in a brainstorming session to organize a conference in the fall of 2004, technology publisher Tim O'Reilly and web pioneer Dale Dougherty noted that the organizations and companies that survived the crash were more important than ever and that they shared similar business models, design, and development patterns. The term *Web 2.0* was adopted to describe the emergent physiognomy of the web and to name the upcoming conference.

The term *Web 2.0* has been criticized as being a marketing buzzword and as promoting the idea of a technological revolution that did not happen, but O'Reilly and Dougherty's analysis behind the term was indeed about the market and the economy, and about a "natural selection" among existing models rather than a *tabula rasa* revolution. To clarify his thoughts, in 2005 O'Reilly published the seminal article "What is Web 2.0".

This entry explores the characteristics drawn by O'Reilly of Web 2.0 phenomenon, how most of them succeeded in changing the way contents and software are produced, how a few others failed, and the societal implications of both of these success and failures. It then discusses if the Semantic Web, sometimes called "Web 3.0", can be considered as the successor of Web 2.0.

From Broadcast to Participation

User-Generated Contents

A common feature can be found in most of the websites that emerged or were strengthened after the dot-com collapse: Facebook, Twitter, YouTube, Amazon, eBay, Wikipedia, the Internet Movie Database (IMDb), or even Internet dating and blogs platforms. All of these sites host valuable user-generated contents. Even Google Search, by ranking sites according to hyperlink topology, collects and processes links created by Internet users. In that way, all those top websites differ from previous ones that were mere translations of corporate communication and mass media broadcast. Rather than producing content, owners focus on providing an “architecture of participation” so that patrons create contents and hence value for them.

Long Tail Economy

A consequence of participation is that it broadens the range of described items and interested patrons. With such a turn, companies have been said to be able to move from a *mass production economy* (focusing on the 20 per cent of products that content 80 per cent of customers) to a *long tail economy* (in reference to the shape of a power law distribution) by offering also the other 80 percent of products, those that can be described as niche products. While it is easy to imagine how this theory applies to e-commerce and even to culture, O'Reilly showed that it applied also to Internet advertising. In the late 1990's, DoubleClick, by serving targeted ads to major websites, succeeded in being ranked in the top 10 websites in the world. However, while very technologically innovative (with early use of web services, mashup and cookies), DoubleClick was limited by its business model: It proudly cited "over 2000 successful implementations" when Google AdSense served hundreds of thousands of

advertisers on virtually any Web page. Interestingly enough, the hindrance caused by DoubleClick's business model was embodied in the use process and the technological architecture since formal contracts had first to be signed with both advertisers and "publishers", to be compared with the simplified process set up by Google AdSense to target niche advertisers and publishers.

Novelty and continuity with the Early Web

With this orientation towards participation, Web 2.0 sites can be easily identified as nearly every page on them can be partially modified by users (typically with an "edit", "comment" or "like" button), resulting in what some called a read/write web. But this is a misleading way of defining Web 2.0 since the web, from its very beginning, was designed with updatability in mind: HTTP, its protocol, includes self-explanatory methods called "POST", "PUT", and "DELETE", and earlier versions included also specialized methods for links and revision management. These methods (with the exception of POST) were nearly never implemented (until the popularization of REST). The reason may be that with no clear vision on who could update what, it was easier for early website owners to rely on a single webmaster who had a direct or FTP access to pages stored as system files.

With the adoption of content management systems (CMS), webmasters gradually disappeared. Pages could be modified by registered users through a web form. While being "read/write", these websites were still not participatory. There is indeed a difference between collaboration and participation. A CMS in which every account registration has to be validated by a manager, in which users are given a "role", in which roles have different access rights on resources, is a *collaboration* platform. By contrast, a *participation* platform is usually characterized by self-registration (with a simple check that the user owns the given e-mail address), a flat hierarchy of roles (at least at registration time, supplemental "powers" being gained by contributing to the community), and often context-dependent rights (e.g. the

right to edit or delete one's own contents). Participation features in Web 2.0 sites make it possible for every user on the Internet to edit any page (or at least comment it, rate it, or create her own and link it). In doing so, Web 2.0 may be one of the best large-scale realization of what the Web was supposed to be: a "hypertext", as defined by Theodor Holm Nelson, the inventor of the concept, in 1965.

Trustfulness and Motivation

User-generated content, along with a model of nearly flat roles, could raise concerns about quality and legality. Two contrasting examples are Wikipedia articles and news on social networks (such as Facebook and Twitter). In its first years, Wikipedia was regarded with suspicion: How could a wiki, editable by anyone on the earth, be as accurate as "real" encyclopedias written by experts? A highly debated study published in *Nature* in 2005 showed that the number of inaccuracies in Wikipedia and in traditional encyclopedias were of the same order of magnitude. Moreover, on Wikipedia, these inaccuracies were fixed soon after publication. Indeed, even critics from the outside are, in a way, a part of the participative process.

One could say that the mantra "With enough eyeballs, all bugs are shallow", first invented for open-source software development, has been shown to fit also content creation, be it through commenting, editing, voting, or abuse reporting. The open-source mantra, however, explains also why participation is not always a guarantee of trustfulness: Contents have to be confronted to enough competing points of view. Massive amounts of fake news spread on social networks during the 2016 United States presidential elections revealed that the virtuous circle of participation can turn into a vicious one when the community is as homogeneous as can be a political activist group.

Another parallel with open-source software has to do with participation incentives: Whereas the altruistic goal of producing common goods is often publicized, the confluence of "selfish" interests into a collective by-product is also a very well attested phenomenon.

Intellectual Property

With user-generated content, the question of users' intellectual property also arises. The default "all rights reserved" copyright notice is difficult to comply with, on platforms reproducing and mixing contents from thousands, millions, or billions of "creators".

A first way was opened by the Creative Commons organization. It transposed open licenses from source code to content, fostering both attribution to creators and free reproduction and remixing. It is probably not a coincidence that Creative Commons was born in the early 2000s, similarly to Web 2.0, and that Wikipedia uses its licenses.

The opposite way is illustrated by participatory websites created prior to Web 2.0, such as IMDb (1990), eBay (1997), NEC CiteSeer (1998). With slightly different phrasings, their terms of service grant intellectual property rights to the content to the publishers of these websites. IMDb's conditions of use, for instance, state that users who post content grant to IMDb a "nonexclusive, royalty-free, perpetual, irrevocable, and fully sublicensable right to use, reproduce, modify, adapt, publish, translate, create derivative works from, distribute, and display such content throughout the world in any media".

With users' increasing awareness of their rights (or lack of rights) on the web, most popular Web 2.0 platforms (e.g., Facebook, YouTube, Twitter, Academia.edu) have dropped the word "perpetual" from their terms of service. In the case of these platforms, the intellectual property license they hold to content posted on the platforms ends when users delete the content or delete their account, with the exception in certain cases of content shared with others who have not deleted it from their own pages.

An alternative way can be found on web platforms targeted at "creators" (Flickr before being sold to Yahoo!, GitHub before 2017, ResearchGate). It consists in requiring no copyright transfer or assignment at all, and letting the creator choose a license for his or her content. In other words, the owners of these platforms that host user-generated content are not considered publishers. Such an interpretation is made possible by the safe harbor provisions in the Online Copyright Infringement Liability Limitation Act, part of the Digital Millennium Copyright Act, and equivalent directives such as the European Union's Electronic Commerce Directive). These provisions shield online service providers from copyright infringement liability on the condition that they take down the infringing content when they are notified.

Sharing Economy or Unregulated Work?

Another social (and possibly legal) aspect of Web 2.0 comes from the fact that several participative platforms can be seen as consumer-to-consumer marketplaces, Platforms such as eBay (for new and secondhand goods but also for services), AirBnB (for short-term lodging), and Uber (for car transportation and delivery) serve as brokers between individual sellers and buyers. While advocated by some as a part of the new sharing economy, this kind of business raises many concerns. Alleging unfair competition (due to licenses, taxes, and other expenses that individuals do not pay), traditional providers of those goods or services regularly protest against individual providers and marketplace owners. In a number of countries, depending on the legality of those marketplaces, governments either try to force recurrent individual providers to comply with some of the regulations and taxes of professional providers, or to take legal actions against the brokers.

In addition to unreported employment, usage of consumer-to-consumer marketplaces could also foster remuneration below minimum wages. On microwork platforms such as Amazon Mechanical Turk, where workers earn a few cents for every "human intelligence

task", estimates of what workers earn range from about \$1.20 to \$5 an hour. Even if most workers are US citizens, as independent contractors, they are not covered by federal minimum-wage legislation.

From Desktop to Web Applications

Rise of Web Applications

Web 2.0 not only affected how content is produced on the web, but also how software is designed, implemented, and managed. As websites became more and more editable, their difference with desktop applications faded. To be precise, web applications were nothing new. As early as 1993, the HTTP protocol was tested by implementing "web mail", a kind of software making it possible to consult and send emails through an HTML page in a web browser rather than through a terminal or desktop client. To stay with this example, 1996 saw the launch of two massively adopted web mail providers: Hotmail (originally styled as HoTMaiL and later acquired by Microsoft) and RocketMail (later acquired by Yahoo!).

Nevertheless, something changed in 2004 with the release of Gmail by Google. The delay between user actions and interface reaction, which was inherent to web applications, seemed to have disappeared. The reason for this was that user actions did not trigger the loading of a new page anymore, but instead the asynchronous loading of structured data (in XML or other format), and, once data are retrieved, the targeted update of the affected HTML interface part. This "new approach to Web applications" was christened "Ajax" by Jesse James Garrett, an expert in web design and user experience, in a 2005 blog post. Technology parts of Ajax—the ability to trigger an asynchronous HTTP request (with XMLHttpRequest and JavaScript), the ability to programmatically modify web page parts (with DOM and JavaScript), and the separation of document content and document presentation (with HTML

and CSS)—were not really new. However these technologies could not have been used so systematically until the end of the so-called browser wars and the incremental fix of browsers incompatibilities.

Changes in Software Engineering

For a very long time, web development was regarded with contempt by software engineers. In the early 2000s, Sun's investment in web development helped to improve its reputation. Owing to Java Enterprise Edition, web technologies were progressively adopted at the core of the information technology architecture of companies, but were also associated with strict yet heavy development and production environments. Web 2.0 changed all of this, with the real assets of software companies becoming data and users rather than software.

First, software quality moved from specifications compliance to customer satisfaction. In order to keep the customer satisfied, as recommended by so-called agile methods, software delivery changed from rare and robust releases to streamlined updates (Flickr was known for releasing bug fixes and new features every half hour, as soon as they were complete). With web applications, centralization of software not only made updates easier, but also made it possible to track users' behavior in real time, to check whether a feature is used or not, to state which special cases cause the process to fail. Iconic Web 2.0 actors, by bearing the beta logo on their applications for years, showed to the masses that they realized with this new release process, that software, by nature, is in an endless cycle of both improvements and non-achievement.

Second, granting value to data and connected users rather than to code could explain a move from software engineers' to system administrators' views on how to develop software. According to O'Reilly, this could explain the move in corporate Web development from compiled languages (e.g., Java) to script languages (e.g., JavaScript, PHP, or Ruby). This

change in "values" could also explain why Web 2.0 development is characterized by extensive reuse of and contribution to open-source software; examples include Node.js package management and the Ruby ecosystem). Interpreting this move as altruistic would probably be naive as the competitive advantage of companies is no longer simply in the source code but in the "service". Software without data and user communities is worth nothing. This emerging idea that source code is disposable could also explain the growing adoption of "code refactoring", the agile process of regularly rewriting code that already works instead of striving to get the most general and efficient code at first.

Third, Ajax, with readable scripts on the client-side and traceable calls to web services, is "hackable" by design, not necessarily in the negative sense of the term (since openness does not contradict security), but in the sense of being open to clever ways of overcoming limitations to get unexpected outcomes. Providing parts of successful web application as "white boxes" (systems whose internals are not hidden) fostered remixability and the creation of value-added services in the form of "mashups".

Distribution or Centralization?

Among the features of Web 2.0 predicted by O'Reilly, distribution was one of the very few that were not corroborated in the following years. In 2005, O'Reilly analysed the success of simple and reusable web services as a forerunner of a more distributed web, a network of cooperating data services both providing and consuming services with others. This could have been the case if the most consumed services were not provided by very small numbers of companies (e.g., Google, Amazon, Facebook, Twitter), which themselves are not dependent on third party services.

The over-optimism of the forecast could be explained by the political nature, among computer scientists, of the debate about centralized vs decentralized computing. Rather than

being a mere technical design choice, modelling relations between computers is considered as setting the relations of power between computer owners. In other words, in such a view, software deployment architecture defines political or economic rules among people (ranging from an authoritarian regime to anarchy, or from a planned economy to a free market). Therefore, it could appear as a paradox that Web 2.0 lead both to a flat participation model among users and to an uneven collaboration model between organizations.

An iconic example of service consumers dependence to service providers is given by a post on Google Code's official blog, in June 2011, entitled "Spring cleaning for some of our APIs", announcing that following the launch of new services, a number of old ones, such as SideWiki and Wave APIs, would be deprecated, or phased out, and already deprecated ones, such as ImageSearch, PatentSearch and Translate APIs, would be shut down). Hundreds of comments were posted in the next few days. Most of them concerned the Translate API, the fact that their applications or services completely depended on it, a lack of good alternatives, the time and effort they invested in these APIs, or the danger of consuming *free* services without paying attention to the lack of guaranties in their terms of service.

This illustration of a very unbalanced market in favour of service providers may be softened: Several commenters tried to negotiate the preservation of the service as a paid one, others argued that the shutdown would cause them to lose their trust and consideration in the provider, and others, as a provocation, mentioned Microsoft's alternatives. A week later, in response to the comments, the post author declared that Google would update its plan with a paid version of the Translate API.

On a more general level, the tendency of Web 2.0 towards an oligopoly can be summarized by the difference in the use of the term API, or application programming interface. In 2005, when O'Reilly spoke about APIs, he was referring to software API, the specifications that were necessary to reuse third party subroutines (especially from the

operating system). According to his paper, control over API was something from the past, from the PC era, in which all the power was owned by operating system creators. As for him, the resulting lock-in (unability to switch to another vendor) would fade away with software services, because the use of open standards would make them interchangeable.

By 2011, API referred to service API, the specifications necessary to reuse third party services. The lack of standard service API and hence the lack of alternative implementations cause the same risk of vendor lock-in as did the "old time" software API. As was the case during the PC era, one key to market control is still the user base, as was illustrated by Google's failed attempt to challenge Facebook with its Google Plus network. But as O'Reilly predicted, the control of data, in the absence of which there cannot be any service, is also key to market control.

Should the Semantic Web be called Web 3.0?

Following the naming pattern of Web 2.0, some have proposed that the Semantic Web (an initiative by the World Wide Web Consortium) could be called Web 3.0. Although no one can predict the future of the web, the term *Web 3.0* would imply that Web 2.0 and the Semantic Web are of the same nature, which is misleading. Web 2.0 is not about a consortium defining standard protocols and formats to enhance an existing infrastructure. Instead, Web 2.0 emerged from independent socio-economic actors. A change in the use patterns of the web and in the way software is engineered, a change that was imperceptible for years, suddenly become dominant.

Another radical difference between the Semantic Web and Web 2.0 is in the way each manages *intelligence*. While the first one aims at "content that is meaningful to computers", the second one aims at "harnessing collective intelligence" from users, as their respective

advocates put it. To be more explicit, in the Semantic Web perspective, documents' meaning should be described with concepts, taken from ontologies specifying the formal links between those concepts. By contrast, in the Web 2.0 perspective, documents' meaning is revealed by readers through comments and tags (free keywords). This rephrasing of the meaning is not only contextual to the document, but also subjective to the readers (with interesting similarities and differences in tags' reuse from one document to the others), and intersubjective (with meaningful consensuses and dissensuses among readers). Besides being of different nature, both approaches have such radically different philosophies and aims that one should consider them as parallel and competing dynamics rather than one being the successor of the other.

Aurélien Béné

See also Crowdsourcing; eBay; Internet Law; Internet Origins and History; Neogeography; Peer-to-Peer Networks; Semantic Web; Sharing Economy and the Internet; Social Media; Webpage Design; Wikipedia

FURTHER READINGS

Allen, M. 2013. What was Web 2.0? Versions as the dominant mode of internet history. *New Media & Society* 15(2):260–275.

Brake, D. 2014. Are we all online content creators now? Web 2.0 and digital divides. *Journal of Computer-Mediated Communication* 19(3): 591–609.

Garrett, J. J., Ajax: A New Approach to Web Applications, Blog post, February 18, 2005. *Adaptive Path*. <http://adaptivepath.org/ideas/ajax-new-approach-web-applications/>

Giles J. 2005. Internet encyclopaedias go head to head, *Nature*, December 15:900–01.

IMDb (n.d.). IMDb Conditions of Use. <http://www.imdb.com/conditions>

Madden, M. and S. Fox. 2006. Riding the waves of “Web 2.0”. Pew Internet Project
http://clearwww.co-bw.com/Ethics/ethics_Web_2.pdf

O'Reilly, T. 2005. What is Web 2.0: Design Patterns and Business Models for the
Next Generation of Software. Social Science Research Network Working Paper Series.