

The Performance of Distributed News Aggregators

Wolf Richter, Tobias Escher, David Bray

Introduction

The production of news has always relied on employing cutting edge communication infrastructure to provide the most up to date information to the readers or later viewers as quickly as possible. But recent advances in digital communication technology have not only enabled a significant improvement in the internal performance of established news organizations, but have also questioned the structure of the industry as such.

With the widespread availability of news-gathering devices, like digital cameras, laptops, and mobile phones, and public access to the global distribution infrastructure of the Internet for most citizens in developed countries, the exclusive power of the news agencies and publishers to collect and disseminate news has received a significant blow. Mobile phone users can take digital pictures with the built-in cameras and distribute the information together with a short story on a blog or another publicly accessible news forum, which turns the mobile user into a Citizen Journalist.¹ The immediate gathering and the direct dissemination of information over the Internet enables a shift in the news production process away from industrial publishing companies to the users, readers, and consumers empowered by shared digital information spaces.

This shift comes in two dimensions: (1) An Internet user has the possibility to gather and publish news items over the Internet and (2) the user can search and aggregate her own news from the abundance of freely available content on the Internet. While the first dimension turns every user into a Citizen Journalist, the second empowers Internet users to become their own news editors.

Guiding questions

The rise of the Internet as a means for news production raises two sets of questions: First, what different modes of production and aggregation have emerged in the online news space and how does their performance compare to each other? Second, do the new paradigms of news production threaten to replace or promise to complement the traditional mode of news production? Or put differently: Is the question to hire a professional news team or to rely on an online news aggregator a "make or buy" decision for a business, or is it a paradigmatic decision with respect to the outcome, i.e. the kind of news produced?

The space for Internet news production is not dominated by one homogeneous production paradigm. On the contrary, a variety of different production paradigms have emerged within the space of online news production, with the Web presence of a

¹ See Dan Gillmor: *We the Media: Grassroots Journalism by the People, for the People*, 2004, Ch. 2.

traditional news production organization like the BBC on one end of the spectrum, and with the decentralized blogosphere, where a majority of amateurs share their thoughts and opinions, on the other end.

The Performance of News production

News is a relative good, because what information is “news” is determined by subjective and geo-temporal factors. We will focus on two characteristics that Galtung and Ruge identified as critical: *Frequency*, i.e. sudden events which concur close to the organization's rhythm of working are more likely to be covered, for example a particular murder case rather than a long-term trend, and *continuity*, i.e. events that are already in the news have a higher chance of being selected as news again.²

The news production market is not supply-constrained. On the contrary, the amount of information gathered and disseminated every day is ever growing and exceeds by far an individual's capacity to process. This poses a challenge for the reader and the news organization to maintain situational awareness in today's global news environment. Even with globally distributed field reporters, the capacity of a single organization may in fact be near the limit of impossibility. For the individual reader, news organizations have so far played an important role as gatekeepers to preprocess and select the news items before publication.³ This function is ambivalent: On the one hand side they help their readers to manage the amount of information they can handle, on the other side they take important decisions on what information readers can access on their behalf. In a digital environment, physical pages or broadcasting minutes are no longer the limiting factor. It is only the bandwidth of the readers or viewers to process, i.e. read or watch the information available at a given day. Therefore also online news production needs to address the challenge of information selection to avoid information overload.

The Nature of the News production problem

The production process consists by and large of two different problems: The first is gathering or reporting of news items and the second is selecting relevant news items for publication. The production process of traditional media production follows “filter, then publish”.⁴ The editorial board of the publisher allocates the limited (physical) bandwidth to certain topics. Only the items, which have been selected to cover a topic, are published. On the Internet, a new production paradigm has appeared: “publish, then filter”. With almost unlimited physical bandwidth, almost every story and every accompanying piece of information can be stored and potentially published in digital format. The role of online news producers is to assist the users to filter the relevant information from all published information on the Internet.

² See J. Galtung and M. H. Ruge: The Structure of Foreign News: The Presentation of the Congo, Cuba and Cyprus Crises in Four Norwegian Newspapers *in* Journal of Peace Research 2(1), p. 64ff (1965); see T. Harcup and D. O'Neill: What Is News? Galtung and Ruge revisited *in* Journalism Studies 2, p. 261ff (2001).

³ See e.g. David Manning White: The 'Gate Keeper' A case study in the selection of news *in* Lewis Anthony Dexter and David Manning White (Publisher): People, society and mass communications (1964).

⁴ See Clay Shirky: The Music Business and the Big Flip *available at* http://www.shirky.com/writings/music_flip.html (last visited 15/11/2007).

Modes of News production

To map the Internet News aggregator space we have decided to choose one representative for each specific mode of production:

Digg News⁵ is a proprietary platform that is open for unedited content submissions: it catalogs news stories submitted by its users as well as questions and comments, and invites users to rate the stories. The content with the most favorable votes is shown on Digg's front page. Digg's agenda is purely determined by the popular vote of its users – the wisdom of the crowd and users have to rely on that what others find relevant or interesting will also catch their interest.

Google News⁶ is a proprietary platform that automatically collects news stories from news sites from all over the world, sorts and tags them using a proprietary classification technology, which makes the news stories searchable using Google's proprietary search technology. In its presentation Google News resembles most closely the look and feel of the online presence of a traditional news producer.

OhMyNews⁷ is a citizen journalism platform that relies on original news contributions written and submitted by its members. Originally from Korea, OhMyNews also publishes an English language edition and has successfully established itself as the global market leader for citizen journalism. Members are volunteering their time. The articles are reviewed and selected by a professional board of editors that enforces the publicized style guidelines and chooses articles by their subjective importance.

bbc.co.uk News⁸ is the online presence of the BBC, which presents headlines and in-depth news coverage created by professional journalists on the Internet. Although bbc.co.uk employs dedicated online editorial staff, which is not integrated into the news production of the other BBC channels, radio and TV, and has adopted elements of citizen journalism like user forums to post comments or pictures taken by users, it is still rooted in the traditional, centrally managed approach to journalism. Therefore we will use bbc.co.uk to represent the traditional alternative to distributed online news production.

The Blogosphere cannot be reduced to one platform, which could represent this paradigm of online news production, which is made up by thousands of individual sources of information, the individual blogs. Blogs can be maintained by individuals or by groups, by amateurs or professionals, and cover all topics the bloggers find worthwhile writing about. We have picked a particular network of blogs, **Global Voices**,⁹ which is dedicated to international news reporting and aggregates a large number of disparate blogs in all parts of the world and should be representative for the mode of production and the range of topics covered in the blogosphere with regards to international news. **Wikipedia**, well known for its peer-produced online encyclopedia, has recently emerged also as a platform to share news.

⁵ <http://digg.com> (last visited 24/11/2007).

⁶ <http://news.google.com> (last visited 24/11/2007).

⁷ <http://english.ohmynews.com> (last visited 24/11/2007).

⁸ <http://news.bbc.co.uk> (last visited 24/11/2007).

⁹ <http://www.globalvoicesonline.org> (last visited 22/11/2007).

Table: A functional analysis of the production functions of news production organizations

News reporting	Production process	News selection	Case studies
Professional reporters create original content	Filter then publish	Editorial Board	Iht.com, bbc.co.uk
Citizen reporters create original content	Filter then publish	Editorial Board	OhMyNews
Automatic algorithm collects content from all kinds of sources	Publish then filter	Algorithm	Google News
Users submit content from other sources	Publish then filter	Public vote	Digg (Reddit, del.icio.us)
Users creates original content or comments on content from other sources	Publish then filter	Social network	Global Voices, Wikipedia/Wikinews

Key findings of the quantitative analysis

Analyzing the RSS feeds of two weeks of news production in early December 2007, we have found that the frequency of edited news producers bbc.co.uk, iht.com, and OhMyNews is determined by clearly identifiable updating cycles: Digg and Global Voices show more continuous update patterns. Their frequency is set by the course of events, not the other way round. All platforms are affected by a reduction of activity over the weekend.

The traditional news publishers changed the articles in the analyzed news feeds more often and covered more topics per period than the analyzed distributed platforms. This somehow counterintuitive finding is highly dependent on the analyzed news feeds. When looking at technology news, the articles on Digg are replaced 4 times faster than on bbc.co.uk and iht.com. This finding leads to the assumption that the differences in continuity between the production paradigms of Digg and the traditional publishers are less structural, but more related to the attention the shapers of the agenda devote to the topic. The technology-affine community at Digg produces technology information more quickly and keeps them more up to date, while technology at the traditional media is only a sideline.

What do distributed news production platforms do "better"?

From our analysis we conclude that the distributed platforms seem to be particularly strong in shaping a news agenda that is in line with their users' interests and thereby expand the range of information available to the general public. Digg's performance in technology news is stronger than that of its non-distributed alternative. Global Voices and OhMyNews have created a news profile, which selects news that would not have been available to the readers through the traditional channels. While the broadening of

the range of news constitutes clearly "better" performance, the 'alignedness' with special interests is controversial: The ethics of traditional journalism have avoided to only satisfy public demand for information. At least in theory, media have a duty to broadly inform the public and raise and keep attention on unpopular, but important topics.¹⁰ Cass Sunstein has raised the concern that the Daily Me created by tapping into the blogs that only reflect the reader's own opinion constituted an "Echo Chamber", which reiterates the point of view of the reader and makes it easier to avoid getting exposed to information the reader does not like.¹¹

It may be that the capacity of traditional news organizations, even with globally distributed field reporters, is not sufficient to cover today's complex news environment. When the campus shootings occurred in the United States at Virginia Tech, Wikipedia actually had the most-up-to-date information compared to traditional news agencies. In addition, several intelligence agencies in the U.S. have recently re-examined their own traditional approaches to gathering intelligence, and in at least one instance a set of experts concluded that single agency-approaches to gathering intelligence were incapable of dealing with the interdependent events of the modern day.

Having reviewed several distributed modes of news production in this paper, we suggest that web technologies may represent viable solutions to address these challenges and represent the next stage of development in the evolution of intelligence and news gathering activities.

Who captures the benefits?

- **Readers** may perceive the pre-selected access to information that reflect their opinion and meet their interests as a benefit, which saves them time to skim through information on topics they are not interested in. On the other side, to avoid the "Echo Chamber" effect, a reader needs to review several sources of information on the Internet to assemble a complete and balanced picture of the news.
- New **intermediaries**, which position their tools of getting access to and selecting information against the traditional media, capture the benefits of drawing significant traffic to their pages. Digg News and OhMyNews monetize the attention by selling ad space on their sites. In addition, OhMyNews is selling added-value information services to paying customers.
- **Bloggers** capture the benefit of increased credibility by being featured on a news aggregator site. The "peer reviewed" of the public vote or an editorial board indicates that they are not alone with their view of the world. We have seen in our study that the peer review can also produce erroneous results, but that after some time errors will be corrected and false information be identified as such through self-correction mechanisms.
- **Traditional news reporters** capture benefits by reviewing news aggregators to find topics, which may have so far evaded their attention. Thereby they increase their network of informants. Traditional media have in recent years reduced the amount of permanent staff reporters. As a result, the range of topics or regions covered by an individual reporter has increased. Also many reporters work as free-lancers and are hired to cover particular stories only. Blogs can act as a point of entry to start an investigation. Unscrupulous reporters will even capture more benefits by misappropriating content originally published in a blog and save the time and effort

¹⁰ See Gillmor *supra* fn.1. at p.5.

¹¹ See Cass Sunstein: Republic.com 2.0.

of further investigation by watching the reaction of others on a news aggregator site.

Conclusion

- Distributed news aggregators provide solutions to the same two problems traditional news publishers have solved so far: News reporting and selection
- Traditional news publishers strike a delicate balance between the interests of the readers / viewers and their professional responsibility. Gillmor has documented that as a result of increasing cost pressure, traditional news producers have already tilted their balance towards the interests of the readers, which seem to favor local and celebrity news.¹²
- Each of the analyzed news aggregators approach this balance in the selection of news differently: While Digg News' public voting mechanisms favors an agenda reflecting the particular interests like technology news and geek trivia, the editorial board at Global Voices tries to build an open space for promoting unheard "voices", and the editorial board at OhMyNews even tries to mirror the continuous news agenda of the traditional media.
- In the distributed news reporting networks studied, the news reporters feel a strong sense of community and identification with the network they contribute to. The embedded case study on the publication of the Blu-Ray code on Digg shows that a platform can even be held hostage by its contributors, who enforce the principles they believe in even against the founders of the platform. The contributors to Global Voices and OhMyNews seem to be motivated by building an alternative source of information, which more closely resembles their perception of the world. The "public good" motivation seems to be a strong unifying force. At the same time, the focus of attention of the contributors is put – following the self-selection principle - more on niche topics, which the contributors deem subjectively relevant and underrepresented in the traditional media.
- Motivation and identity in all aggregators with human contribution, stems from complementing or adjusting the agenda set by the mainstream, read traditional media. The agendas do not seem to be totally detached, but rather complementary as many topics brought up in Digg or commented on in the blogosphere relate to content produced by traditional media. The ratings and comments can direct attention to underrepresented topics or reevaluate the opinions expressed by the traditional media.
- Therefore, it seems highly unlikely that the news production industry can currently think of using distributed news producing networks in a "make or buy"-decision manner.
- At the same we have a strong hypothesis that the traditional news organization is no longer capable of coping with the complexity of covering the full range of global events and that distributed news gathering is a viable way going forward.
- Popular voting mechanisms for news selection seem biased towards particular interest topics and prone to errors, although a strong community like Digg is

¹² See Gillmor *supra* fn.1 at p.5.

capable of correcting mistakes after some time. Still, popular votes do not seem like a reliable mechanism traditional news publishers could rely on to solve the news selection problem. It also seems unlikely that the sense of community required would emerge if a traditional news publisher would attempt to create one.

- The question for traditional media platforms is therefore less how to substitute their traditional model with a distributed model, but how to open their existing model to integrate elements of distributed news gathering and citizen journalism into the news production process. Given the experience of the editorial boards, it seems plausible to leverage the power of Citizen Journalism and other forms of distributed news gathering for the solution of the reporting problem without compromising on the quality standards and professional journalistic responsibility.
- It is too early to estimate the market effect on traditional news publishers. So far, it looks like distributed news aggregators are acting not as a substitute but as a complement, which in one way or the other reference or reject the agenda and the material published by traditional news publishers and hence cannot (yet) exist as an information space on their own.¹³

¹³ Further insights into the interaction between distributed news aggregators and traditional media portals could be derived from the Webmetric method developed by Robert Ackland et. al, see R. Ackland, M. O'Neil, R. Standish. and M. Buchhorn: VOSON: A Web Services Approach for Facilitating Research into Online Networks, paper presented at the Second International Conference on e-Social Science, 28-30 June 2006, University of Manchester *available at* <http://voson.anu.edu.au/papers/ncess-conf06-full-paper-final.pdf> (last visited 20/11/2007).