

Collective Intelligence and the Creation, Use and Management of Citizen-Led Journalism

Tommaso Colombino, Antonietta Grasso and David Martin
Xerox Research Centre Europe, Grenoble, France
Email: {firstname.lastname}@xrce.xerox.com

Introduction

The Internet has provided a new communication infrastructure which offers the public the opportunity to participate in the production of content – content which is then available as, or as a part of, many different kinds of offerings, services, products and resources. The interest and usefulness of the outcome of networked and collaborative activities is amply demonstrated in existing Web 2.0 content (the ever expanding collection of wiki-type resources and social networks being prime examples). The integration of these types of collaborative tools and resources within business organizational processes has been, to date, less than successful. This may have to do with the difficulty for users to integrate stigmergic knowledge and creative practices (i.e. to read, understand and build upon the traces of others' activity) within the relatively inflexible requirements of organizational procedures and processes.

In this position paper we would like to encourage reflection on the opportunity for collective intelligence tools to position themselves across the organizational boundary that generally exists between the creation/production and the consumption of a product, in a domain where the opportunity for an end user to actively collaborate in a creative organizational process may be of genuine value: the production and consumption of 'professional' news.

The editorial process and the potential for collective intelligence

The process through which news is created and the traditional business models which support editorial work are shifting and moving away from a clear distinction between readers and writers, and audiences are no longer gathered around established media and publishing channels. The *life-cycle* of a piece of news can now be in perennial evolution, and processes of professional content creation are under pressure to change, moving from delivering to a mass to being engaged in conversations with the readers. Users are moving from solely providing opinions and feedback on news stories in on-line blogs and comment sections to actively participating in their construction. Readers are becoming more participative in the news creation process, but the degree of their involvement can vary according to many factors. In fact the definition of citizen journalist may encompass a wide range of levels of "engagement" with the news, from those who simply want to be well informed, to those who wish to contribute with their own opinions and snippets of information, to those who actively participate in maintaining and publishing information resources on the internet.

Content provided by non professional users is already finding its way into prime time news. For example, some of the most powerful images from major events such as the attack on the London transport network in July 2005, the storms that swept the UK in January 2007 and the

Asian tsunami of December 2004, were provided by eye-witnesses. This new type of journalism, facilitated in many ways by the widespread availability of photographic and video recording on devices like mobile phones, has been named citizen or sometimes networked journalism. Social network services such as Twitter or Facebook are also huge enablers of this kind of behaviour. Multi-modal information can be shared and/or published from a mobile phone and published instantly with global reach and multimedia support. .

A recent example of the complexity that characterizes the creation and consumption of news on the internet comes from the Barack Obama citizenship conspiracy theories. It is an interesting case because, putting aside the merit of the theories themselves, it was by and large constructed, and ultimately deconstructed, on the Internet through a sequence of what could be defined documentary investigations around a fact that was called into question (Barack Obama's status as a natural born citizen of the United States of America), and the evidence (documentary, anecdotal) that people on both sides of the debate were able to retrieve, analyze and re-distribute over the web. Anyone interested in making sense of this piece of news, or piecing together a version of the event (in this case, the fact that there are people who have called into question the legitimacy of Barack Obama's citizenship, and therefore his eligibility to the position of President) that is supported by reliable evidence, has to undertake a substantial amount of investigatory work:

- understanding the legal definition of "natural born citizen" and on what grounds it can be challenged;
- understanding the (mainly user generated) challenges to Barack Obama's citizenship (place of birth, status of father, etc.), who produced them and what evidence is presented to support them;
- what is the nature of the challenges (technical, legal) and what is the expertise of those presenting evidence to support them;
- what is the standard procedure for establishing eligibility in public office and was that procedure adhered to in this case;
- have similar cases occurred before;

On top of all this a news story is far from being a static affair, and as the data (documents, user generated content, news imagery, footage and textual articles in professional and amateur datasets) is revealed, and gradually becomes available, it is important to be able to keep track of contributions. Also to cross-reference and establish temporal links between events and revelations, and the information and documentary evidence that pertains to them.

The professional press is only just beginning to explore ways of integrating readers in their content creation workflows. Some newspapers, like the New York Times, are exposing their databases of news under a programmatic environment, on top of which new services, tools, visualizations and so on can be built by third parties. Others, like The Guardian, are experimenting with other means of participation like crowdsourcing, by asking the readers to help with little pieces of investigation that may contribute to mainstream news topics. On the other side, many newspapers are trying to experiment with new types of participative journalism and are struggling to figure how to make integrate user generated content in their professional activities and use it to increase their profitability.

There are scenarios where user contributions, and more specifically eye-witness contributions, are, in fact, an essential source for certain types of content. The 2009 Iranian presidential elections provide an interesting case in point. There were active efforts on the part of the Iranian government to block communications and consequently, doubts about the objectivity of the information coming from the official reporting channels. In this situation, eye-witness accounts, and in particular videos and pictures taken with personal devices (such as mobile phones) were not just integrated into news production, but were indispensable in providing alternative, and in the eyes of many people more credible, information about key events, such as the scale of the protests, the behaviour of the authorities, the presence or absence of opposition candidates in the protests and their behaviour and so on. Similar activities have occurred during the recent and on-going problems in Myanmar.

A more recent example comes from the recent protests that took place in Madrid organized by anti-abortion groups against the incumbent government and Prime Minister Zapatero. A point of contention, one that emerges consistently in the reporting of large demonstrations, is that of how many people actually participated in the protests. The organizers provide their numbers, while other parties and the authorities often provide very different ones, where clearly there is a vested interest for all people involved in defining whether the protest was successful or consequential. In this instance the Spanish newspaper El Pais paid a company that uses proprietary software to estimate the number of people present using video and still images. The technology itself will, most likely, become freely available – what will remain paramount is access to media (video and images) of sufficient quality and quantity (and taken from the all the appropriate angles, at a sufficient number of time intervals, etc.). This is exactly the type of contribution that a news organization, given the right tools, infrastructure and production workflow, will want to actively elicit from its readers.

To give another example, during the spring of 2009 a political scandal hit the United Kingdom, when some of the members of the Parliament were discovered issuing fraudulent or ethically questionable reimbursement requests. The amount of available data (expense requests and justifying receipts) was in the order of half a million of documents, and clearly difficult to manage without a substantial investment in time and resources. In this context the Guardian organized a unique experiment: given that the requests and the corresponding receipts had been made public, the newspaper organized a Web site where the documents were made accessible, and asked readers to help categorize and do a first degree of assessment of their potential relevance. A substantial piece of investigative work of this type involves many steps:

- Digitizing the content not already digital and redacting them;
- Classifying the documents (e.g. being a claim or a receipts or something else);
- Assessing a first level of potential interest;
- Making detailed comments and explanations about the suspicious content;

As with the previous examples, this story is far from being a static affair, and as the data is revealed and gradually becomes available it is important to be able to keep track of contributions, and cross-reference and establish temporal links between events and revelations, and the information and documentary evidence that pertains to them.

Conclusion

These examples show that participation in news construction can go far beyond “have your say” unstructured comments or simple collection of photos and videos about an event, but actually become a structural part of the content. This approach, however, requires support for the readers to understand how their contribution can play a role and support on the side of the editorial editor to analyse and embed user generated content in an evolving news line. To make it viable for the press to offer news as evolving participatory stories, there are high level technological challenges which need to be addressed:

- providing **custom** news provision that supports the reader in understanding of the unfolding of the story and how and where user input can further contribute to the evolution of it;
- **managing development** of news stories. Not necessarily controlling but enabling some transparency of sourcing, paths of development, use of opinion and so forth.
- analysing and **making sense** of large volumes of user generated content and feeding them back in the story
- **assembling** diverse (in terms of media, source, quality, perspective, etc.) content in large volumes into coherent items of news;

These in turn may involve a number of more concrete challenges that directly involve technologies relevant to CI applications and content that can range from unstructured and textual to fully multimodal (images, sound, video, text):

:

- Aggregate access to the various posts and articles on a topic;
- Highlight relevant entities and the concepts that explain them;
- Support to guide the redaction phase prior to publish the documents, when privacy requires so;
- Automatic categorization based on a subset of annotated data collected from the newspaper readers;
- Quality assessment measures to detect cases of mistakes from the annotation and filtering collected.
- Visualizations of e.g. development of the reasoning along time, authors of contributions and degree of trust that could be associated to them.
- Technology that traces the roots of source material and its development and transformation.

References

Participatory initiative by The Guardian: <http://mps-expenses.guardian.co.uk/>

User generated news resources: <http://www.nowpublic.com/> , <http://news.yahoo.com/you-witness-news>

Sabadin, Vittorio. The Last Issue of the New York Times : The Future of Newspapers. Editorial Solgo Media, 2007.