

Crowdsourcing and Gov 2.0

Irene Ros, Yannick Assogba, Joan DiMicco

IBM Research

One Rogers St, Cambridge, 02142

[iros; yannick; joan.dimicco]@us.ibm.com

INTRODUCTION

The openness and proliferation of the social web inspired a movement known as Gov 2.0 or Open Government. The Gov 2.0 movement strives to make our government as accessible, transparent and collaborative as the services being used by Internet users on a daily basis. There are many definitions of Gov 2.0 [3]; Some focus on real time data availability and some argue the focus should be on using the principles exhibited by the emerging tools & technology in the social media space. We see at least three different strands in the work coming out of the Gov 2.0 movement: informing citizens, gathering citizen input or encouraging citizen participation and providing better services to citizens. This position paper intends to outline aspects of the Gov 2.0 movement, with particular reference to the software systems that are part of this ecosystem, as well as our view of how these can be impacted by crowdsourcing techniques.

The community of Gov 2.0 thinkers, spearheaded by Tim O'Reilly, faces several challenges as a result of the government status quo and nature of existing operations. The digitized government, still a novel concept, is one of great cost and labor. Among many factors, accessibility requirements, bidding processes, reliability standards and a wide target demographic result in expensive investments that take months and even years to make. Additionally, many argue that the government itself cannot fund and lead the movement of its own openness; that only an outsider can function as a "watchdog", supervising the activity and data generated by our government representatives. Despite the challenges, President Obama's memorandum on Open Government [10] called on the US government to become more transparent and open.

THE OPEN GOVERNMENT DATA AND SOCIAL WEB MOVEMENTS

Many government offices answered the President's call, releasing copious amount of data. This data, often available as single file dumps, or worse - html only interactive tables, are difficult to turn into actual insight. What are the ramifications of this data flood? First, it offers the illusion that the government is now transparent and its operations can be easily discerned from the data. This is far from the truth: citizens are not equipped or motivated to parse large data dumps and build their conclusions upon it. Much of the data being released is of little importance to the majority of citizens and some have suggested that this data is released merely for the sake of appearing transparent [4]. These

issues alongside those mentioned above motivate a community of developers who wish to convert the data into meaningful interactions for citizens, as indicated by examples such as "Code for America" [5] and "Apps for America" [2]. Many applications are being built on top of data, some utilizing concepts we know from the social web such as commenting, sharing, ranking, embedding and remixing. A prominent example being OpenCongress.org which offers a robust commenting system on the contents of legislative bills. Are the sorts of applications being created today sufficient for meeting the goals of the gov 2.0 movement? How can, and does crowdsourcing play a role in the gov 2.0 ecosystem?

GENRES OF CROWDSOURCING AND THEIR APPLICABILITY TO THE OPEN GOVERNMENT SPACE

Quinn and Bederson define a taxonomy of distributed human labor genres and applications [11]. These include:

- Games with a purpose: Applications that employ game mechanics to encourage volunteers to perform computation. The focus is to make the game fun enough that volunteers enjoy playing it while performing work as a side effect.
- Mechanized Labor: Crowdsourcing applications that involved monetary rewards for human labor. The most notable example being Amazon's Mechanical Turk [1].
- Wisdom of Crowds: The collective intelligence obtained by a distributed group of people thinking independently. [12]
- Dual-purpose Work: This type of work is characterized by coupling computation with something else that the user is required to do, an example of this is ReCaptcha [13].
- Grand Search: Work that utilizes volunteers to search through a large data space to retrieve a single solution.
- Human-based Genetic Algorithms: An approach where the solutions consist of a series of small tasks that evolve based on human evaluation.
- Knowledge Collection from Volunteer Contributors - This is using human workers to build large datasets on a common topic.

While the space of approaches is large, some are more difficult to apply to the Gov 2.0 space successfully due to the nature of the work in question. For example, creating a

game around government data that offers sufficient entertainment to satisfy players is quite challenging. In the case of mechanized labor, which can be applied to almost any situation where the task can be split up efficiently, when it comes to understanding the operations of the government, some tasks require prior knowledge and context to be executed accurately. Given that open government work is often not a revenue generating business, budgeting for mechanized labor is another challenge. Lastly, tasks that aggregate the wisdom of the crowd to create a knowledge set or rely on volunteers may be at risk of inaccuracy due to contamination by political bias. Citizens participating in online political activities chose to do so to promote their view points; when collecting information from these citizens, the resulting data can be skewed by position or political leanings. When obtaining input from citizens about government decisions and operation, this form of input may be very valuable, but when attempting to derive a collective knowledge that is impartial, it can hamper one's effort. A familiar example of this is Wikipedia's edit wars on controversial topics.

There are some success stories in using crowdsourcing in the context of Open Government. While monetary rewards and fun may be harder to offer, due to the nature of the work, altruism can play a substantial role when the work is positioned to contribute to the greater good. A famous example of such crowdsourcing is the Guardian's "Investigate your MP's expenses" applications [6]. In this tool, users can review the expenses of their Members of Parliament by marking suspicious parts for review and at the same time digitizing them. This example of volunteer-based crowdsourcing allows citizens to rally in protest against unreasonable government spending by performing their small share of work. The application has been tremendously successful, reviewing over 200,000 documents and finding numerous outrageous news-worthy examples.

While the task given to individuals by the Guardian is fairly small, there are thousands of documents to digitize. We suggest that there is space to explore crowdsourcing on a different level, where the tasks are large and the participants are few. An example of this is seen in the Massachusetts Bay Transportation Authority's (MBTA) approach to making use of their real-time data on the public transport system. Given budgetary constraints, the MBTA could not build the web and mobile applications fully utilizing their real-time train data. Instead, the MBTA released the real-time tracking data and spearheaded a grassroots effort to encourage independent developers to build a variety of applications around the data; effectively crowdsourcing the development of services to deliver to the public. This resulted in making the data useful to the larger public through many applications and keeping the cost low for the MBTA and taxpayers.

CROWDSOURCING ANOMALIES IN CONGRESSIONAL LEGISLATION

We are currently exploring crowdsourcing opportunities in our own work in the Gov 2.0 area, in a project that visualizes visualizing US Congressional Legislation: IBM Many Bills (<http://manybills.us>). Our initial deployment of Many Bills focused on offering an easy interface to reading legislation [9]. We focused on the use of color and abstraction levels to offer varying degrees of bills' content: the lowest level being text-free colored boxes representing individual sections, with the color signifying the section's topic, and the highest level of detail being the full text of a bill's section. Early on in the project, one of our goals was to make it easier to identify outliers or off-topic "riders" in bills. These are sections, that due to the vagaries of the political process, end up attached to bills that they are topically unrelated to. A prominent example of this is the Credit CARD Act of 2009. The purpose of this bill was to introduce legislation that protects consumers from predatory practices by credit card companies. While the majority of the bill focuses on the topic, the penultimate section doesn't—this section introduced into law the right of Americans to bear automatic weapons in US National Parks.

While this example clearly demonstrates an outlier within a bill, more often than not, off-topic sections are harder to detect. From the beginning, it was our goal to devise an algorithm that would detect such outliers sections automatically. After many attempts, our recent work employing natural language processing techniques allowed us to find candidate outliers in our corpus. We quickly realized that human judgement on the validity of these ratings is far superior to what our algorithms could produce through automated techniques alone.

To bring in human judgement to validate the machine ratings, we considered using a mechanized system of distributed work such as Amazon Mechanical Turk and offering payment. Because the work of judging bill content is both time-consuming and highly intellectual (requiring reading and understanding an entire bill), yet also requires minimal feedback from the user ('yes' if the section is an outlier, 'no' if it isn't), we felt we had little chance of ensuring quality of work if left to anonymous, paid users.

Instead, we decided to add a crowdsourcing feature to the site that would allow users on the site to vote on whether a section is an outlier or not by showing users our automatically-detected anomalies and asking for their judgement, on a volunteer basis. Users can flag any section as an outlier, whether or not it has been flagged as an anomaly -- the machine results act as a guide to potentially-promising parts of the dataset. The system uses a voting mechanism to tally votes from the volunteer users on each bill section that has multiple votes, and displays the user-determined outlier status of the bill, overriding the machine-generated judgement.

Within Quinn and Bederson's taxonomy, this is a volunteer-based, grand-search effort relying on altruism. The task will benefit most from participants who have topical knowledge,

legislative text understanding and desire to read legislation. The feature has been built and is just about to be deployed on the site -- we are still unsure as to whether altruism will be a strong enough motivator to encourage participation. To further encourage users to participate in the crowdsourcing activity, we have built different users ranking features, such as a display of the top outlier judgement contributors and awarding small badges on users' profile pages if they are top contributors. We are unclear whether these forms of reward will be enough additional incentive for users to contribute, and this will be the main focus of our next research on the site.

FUTURE WORK

The example given above of the MBTA shows us that by unlocking the creativity of relatively small numbers of highly-skilled individuals, raw government data can be transformed into useful information that citizens can act upon. This layer of translators will be critical to the success of the Open Government movement and could possibly benefit from increased capability to incorporate crowdsourcing techniques in their work. What would an appropriate crowdsourcing platform look like for them? If the task is not easily amenable to a micropayment piecework system like Mechanical Turk, would it be possible to build a platform that attracts and retains interested volunteers to do piecework in the name of open government? These workers would likely be more motivated and capable of doing difficult tasks.

An example of something that would live in this ecosystem is ScraperWiki [8] a wiki-like system for screen scrapers of government data. Volunteers upload code for scrapers and modify existing ones to keep them working or add functionality. Additionally ScraperWiki provides the infrastructure to actually run the code and store the data, taking the hosting burden off the volunteers. Could mechanisms like this be extended to the wholesale building of applications and the data processing needed to drive them? We suggest that this community of mediators provide an opportunity for useful research into applications of crowdsourcing in this domain.

We have mentioned issues of political bias when it comes to crowd participation in Gov 2.0 endeavours. While this is not always a bad thing, one could conjecture that, in a system like Many Bills, showing users the breakdown of political sponsorship of a bill (Republican vs Democrat) might influence whether a user is likely to mark a section within a bill as an outlier or not. More generally, research into how the presentation of crowdsourced tasks can bias the quality of responses could be beneficial for the community of open government developers, who are constructing applications that depend on crowdsourced labor.

We assembled a small group of volunteers (including ourselves) in an attempt to generate a ground truth dataset of outlier sections for Many Bills. We soon realized that the judgement of whether a section is an outlier or not can be

subjective and may be affected by our individual political biases. Reaching an agreement among the group members required extensive discussion and clarification of the text. This type of operation, an intermediate step of negotiating between workers is not currently available in typical mechanized labor applications (though it is present in other genres of crowdsourced work). We wonder if it may prove beneficial for complex task, when collaboration, rather than duplication, is desirable. Research into whether this could successfully be integrated into typical mechanized labor setups may prove interesting.

CONCLUSION

The Open Government movement is a recent endeavor seeking to better inform citizens, collect input from citizens and offer better services to citizens. While many applications have been built in this space, few utilize the crowd in one of the many ways outlined in Quinn and Bederson's taxonomy. Our survey of the space leads us to believe that focusing on skilled workers is a necessity for tasks in the Open Government space, due to the complexity of the data. Given this, intrinsic motivations (altruism) are more likely to produce quality data than monetary rewards alone (particularly when payments are small). Our own implementation of a crowdsourcing feature in Many Bills is exploring this issue starting from the use of intrinsic motivation. We believe several threads of future work emerge from our overview: ways to better enable skilled workers in smaller communities of volunteer labor (such as the independent developer communities engaged by the MBTA) to use crowdsourcing could empower them to create more powerful applications. If Open Government crowdsourcing efforts may suffer from worker's political biases, research into what kind of information presentation increases this risk and how to best avoid it could be helpful to the Gov 2.0 community. Lastly, given the complexity of government data, exploring the utility of an intermediate negotiation step between crowd workers themselves could be beneficial to the field.

REFERENCES

1. Amazon Mechanical Turk.
<https://www.mturk.com/mturk/welcome>.
2. Apps for America.
<http://sunlightlabs.com/contests/appsforamerica/>.
3. What is Gov 2.0 Video: Gov 2.0 Summit - Co-produced by TechWeb & O'Reilly Conferences, September 09 - 10, 2009, Washington, DC. 2011.
<http://www.gov2summit.com/gov2009/public/content/gov2-video>.
4. YouTube - Gov 2.0 Summit 2010: Ellen Miller, "Open Government Scorecard". 2011.
<http://www.youtube.com/watch?v=UNQteT9Bu2w>.
5. Code for America. 2011. <http://codeforamerica.org/>.
6. Investigate your MP's expenses. 2011.
<http://mps-expenses.guardian.co.uk/>.

7. H.R.627 -- Credit CARD Act of 2009. The Library of Congress, THOMAS, 2011.
<http://thomas.loc.gov/cgi-bin/query/D?c111:6:/temp/~c111Y2llwm::>
8. ScraperWiki. 2011. <http://scraperwiki.com/>.
9. Assogba, Y., Ros, I., DiMicco, J., and McKeon, M. Many Bills: Engaging Citizens through Visualizations of Congressional Legislation. To appear in proceedings of the 29th international conference on Human factors in computing systems.
10. Obama, B. Memorandum for the Heads of Executive Departments and Agencies: Transparency and Open Government, The White House. 2011.
http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment/.
11. Quinn, A.J. and Bederson, B.B. A taxonomy of distributed human computation. (2009).
12. Surowiecki, J., Silverman, M.P., and others. The wisdom of crowds. American Journal of Physics 75, (2007), 190.
13. Von Ahn, L., Maurer, B., McMillen, C., Abraham, D., and Blum, M. recaptcha: Human-based character recognition via web security measures. Science 321, 5895 (2008), 1465.

AUTHOR BIOGRAPHIES

Irene Ros is a Research Developer at IBM, in the Visual Communication Lab. She has worked on visualizing congressional legislation (IBM Many Bills), visualizing MBTA traffic patterns (A Day of MBTA) and empowering data visualization amateurs to create visualizations of their data (IBM Many Eyes.) Her interests include building engaging visualizations and user interfaces, exploring ways to understand and represent large datasets and cutting edge web technologies.

Yannick Assogba is a Research Developer in the Visual Communication Lab at IBM Research's Center for Social Software. His research interests include data visualization, networked collective and mediated collaboration.

Joan DiMicco is a Research Scientist at IBM, heading the Visual Communication Lab. Her research interests include social networks, visualization, and collaboration technologies.