Why evaluation isn't a party at the end: Evaluating crowdsourcing websites

Presentation: National Digital Forum conference

27 November 2013, Wellington, New Zealand

Donelle McKinley

PhD candidate School of Information Management

Victoria University of Wellington

donelle.mckinley@vuw.ac.nz

www.digitalglam.org

@donellemckinley

1. Introduction (title)

I'm a PhD candidate in the School of Information Management at Victoria University. My research focuses on non-profit crowdsourcing, and user interface design in particular.

There are two things I'm going to talk about this afternoon. I'm going to show you why evaluating crowdsourcing websites early in the design process can be worthwhile, and I'm going to talk about one method of evaluation that is inexpensive and relatively easy to use.

First I must give credit to David Ellis & Co. (2008), who gave a presentation at Museums and the Web a few years back, and inspired the title of my talk.

2. Crowdsourcing (The Building Inspector home)

Crowdsourcing outsources tasks traditionally performed by specific individuals to a group of people or community through an open call. *Crowdsourcing* author Jeff Howe (2008, p. 280) explains that crowdsourcing isn't a single strategy, but "an umbrella term for a highly varied group of approaches".

Crowdsourcers working in cultural heritage and academic institutions share two common project goals: to create or enhance digitised data, and engage the wider community (Carletti, Giannachi, Price, & McAuley, 2013; Owens, 2013; Wiggins & Crowston, 2010). Volunteers are performing a wide range of tasks, including tagging, text correction, transcription, contextualisation, classification, curation, data collection, description, encoding, mapping, geo-referencing, identification, interpretation and translation (Bonney et al., 2009; Dunn & Hedges, 2012; Oomen & Aroyo, 2011).

This is a new project driven by the clever people at New York Public Library called Building Inspector, which invites volunteers to assist with cleaning up historical map data.[1]

The success of non-profit crowdsourcing initiatives relies on meeting two key objectives: sufficient participation and quality contribution (Oomen & Aroyo, 2011; Simula, 2013; Vukovic, Kumara, & Greenshpan, 2010). Meeting these objectives requires an understanding of contextual factors such as volunteer motivation, as well as effective project and system design, and evaluation and refinements to achieve optimal performance (Brabham, 2013).

---

[1] http://buildinginspector.nypl.org/

3. Website design and evaluation (Linked Jazz home)

A typical website for non-profit crowdsourcing includes a branded homepage that describes the project and invites volunteer participation, and webpages for instructing volunteers and performing tasks. Additional webpages are commonly used for registering volunteers, presenting detailed information about the project and the project team, updating project progress, volunteer accounts and profiles, communication, and presenting the outcome of the project for public use.

This is another recent project called Linked Jazz, driven by the Pratt Institute School of Library Information Science. Volunteers classify relationships between jazz musicians to contribute to a linked open data resource.[2]

Usability author Steve Krug (2006, p. 162) talks about the "reservoir of goodwill" that visitors bring to a website, and explains that each problem they encounter lowers its level. The role of evaluation in the design and optimization of websites is to identify real and potential problems so that they can be remedied.

Usability is a broad concept that refers to "how easy it is for users to learn a system, how efficiently they can use it once they have learned it, and how pleasant it is to use" (Nielsen & Mack, 1994, p. 3). Some of the website elements that impact on usability include content, language, readability, website navigation, arrangement of page elements, consistency, visual appearance, page load speed, and the number and complexity of processes to complete the desired action.

In crowdsourcing terms, these design decisions can impact on the effectiveness of the invitation to participate, task instructions, and incentives; the length of time participants spend on the site; efficient task completion; and participant return rate.

Evaluating early in the design process can help to better meet the needs of your volunteers, and avoid major website fixes later on. I'll give you an example.

---

[2] http://linkedjazz.org/52ndStreet/

McKinley, D. (2013). *Why evaluation isn't a party at the end: Evaluating crowdsourcing websites* [presentation]. National Digital Forum Conference, 27 November 2013, Wellington, New Zealand. Retrieved from http://www.digitalglam.org/crowdsourcing/talks/

4. NZRED (NZ-RED blog)

I'm involved with the New Zealand Reading Experience Database (NZ-RED), which is a crowdsourced history of reading project being developed at Victoria University of Wellington.[3]

Based on a UK project launched in 1996, the NZ-RED will collect reading experiences of New Zealanders from the nineteenth century to the present day.[4] Volunteers will be invited to identify instances of reading in diaries, letters, biographies and memoirs, from private collections, libraries and archives, and contribute their discoveries to the online database.

5. UK-RED home

The fastest and easiest way to develop a user interface for the NZ-RED would be to use the existing UK-RED as a template. But our team had some issues with this.

6. UK-RED task interface

The task interface that UK-RED volunteers use to input reading experiences is a very lengthy, rather daunting, one-page online form.

The current UK-RED website hasn't been subjected to usability testing, and no requirements documentation is available. This raised the question, "How effectively and efficiently is the UK-RED task interface supporting rich data collection and volunteer participation?". Rather than blindly adapt the UK-RED template for our purposes and hope for the best, I was keen to pursue this question.

Earlier this year, as part of the Master of Information Studies programme at Victoria, I conducted a research project on the NZ-RED. The aim of the project was to produce high-level functionality and usability requirements for a NZ-RED task interface, and determine the extent to which the UK-RED met these requirements. The findings will inform the design of a working prototype to be developed in the next stage of the NZ-RED project, which will undergo user testing before being incorporated into the overall NZ-RED website design.

---

[3] http://nzredblog.wordpress.com/
[4] http://www.open.ac.uk/Arts/reading/UK/

McKinley, D. (2013). *Why evaluation isn't a party at the end: Evaluating crowdsourcing websites* [presentation]. National Digital Forum Conference, 27 November 2013, Wellington, New Zealand. Retrieved from http://www.digitalglam.org/crowdsourcing/talks/

I employed several data collection techniques to produce the requirements, but today I'm just focusing on the evaluation of the UK-RED task interface.

7. Heuristics

One of the tools I used to evaluate the usability of the UK-RED task interface was a set of heuristics, or design principles (Rogers, Preece, & Sharp, 2011). These are short statements about what a system should do, which are generally accompanied by more detailed explanations.

Heuristic evaluation is known as a "discount" evaluation method, because it does not involve end users, which is time-consuming and can incur costs, and requires minimal time and resources (Cockton, Woolrych, Hornbaek, & Frokjaer, 2012).

Guided by a set of heuristics, evaluators determine the extent to which user-interface elements comply with design principles. Any breach of heuristics is commonly recorded and rated using a four-level severity scale developed by Molich and Nielsen (Petrie & Power, 2012, p. 2107).

The set of heuristics I used was recently developed by Petrie and Power (2012) to support the design and evaluation of highly interactive websites, such as those requiring users to input information.

An example of these heuristics is "Avoid duplication/excessive effort by users", which is accompanied by the explanation: "Do not ask users to provide the same information more than once and do not ask for excessive effort when this could be achieved more efficiently by the system."

Guided by the twenty-one heuristics, I was able to identify potential difficulties associated with the UK-RED task interface that might impact on the user experience, and report how these might be remedied.

McKinley, D. (2013). *Why evaluation isn't a party at the end: Evaluating crowdsourcing websites* [presentation]. National Digital Forum Conference, 27 November 2013, Wellington, New Zealand. Retrieved from http://www.digitalglam.org/crowdsourcing/talks/

8. HE results

I identified thirty-two potential usability problems, related to interactivity, content, physical presentation and information architecture. Of these, 6 were rated major (high priority), 23 minor (low priority), and 3 cosmetic only.

To validate the results of my heuristic evaluation, I conducted an online survey of current UK-RED volunteers and potential NZ-RED volunteers, who were asked to identify usability problems associated with the UK-RED interface.

9. Survey results

Survey respondents identified a total of 23 problems. Of these, 13 were rated major (high priority), 9 minor (low priority), and 1 cosmetic only. Consistent with the heuristic evaluation, major problems related to Physical Presentation, Content and Interactivity, and most problems related to interactivity.

10. Requirements

The outcome of my research was seven functionality and usability requirements for a NZ-RED task interface that supports volunteer participation and quality contribution: minimize user effort; support integration of the task with research processes; enable new visitors and contributors to understand what the task involves quickly and easily; support accurate and controlled data entry; be easy to use for people reasonably confident with the Web; support flexible, structured data entry; and support bilingual data entry.

11. Takeaways

My research found that the UK-RED task interface only partially meets four of these seven requirements. Its limitations are partially symptomatic of its age; website design and user expectations have evolved since the UK-RED was last redesigned in 2009.

The first takeaway from my project was that using an existing crowdsourcing project template may not effectively serve the needs of your volunteers or your project objectives. The only way you can determine that is by subjecting it to evaluation.

McKinley, D. (2013). *Why evaluation isn't a party at the end: Evaluating crowdsourcing websites* [presentation]. National Digital Forum Conference, 27 November 2013, Wellington, New Zealand. Retrieved from http://www.digitalglam.org/crowdsourcing/talks/

The second takeaway was that heuristics can be an efficient and effective method of evaluation.

12. Future research: user interface design heuristics for non-profit crowdsourcing

As non-profit crowdsourcing websites are highly interactive websites that require users to input information, Petrie and Power's heuristics could be used to support the design and evaluation of other crowdsourcing user interfaces. However, they do have some limitations. Firstly, they were developed using a sample of government websites that users would most likely visit out of necessity rather than choice, and possibly only once.

Secondly, the set of heuristics underplays several key elements that impact on volunteer participation, such as value proposition, motivation and incentive. Thirdly, they don't incorporate relevant crowdsourcing research, such as the tips for crowdsourcing published by Rose Holley (2010) and the guidelines published by Michael Lascarides (2012).

The aim of my PhD research is to develop a set of heuristics for non-profit crowdsourcing that supports user interface design and evaluation practice. The initial set of heuristics will draw on Petrie and Power's heuristics and other relevant sources, and these will be evaluated and refined over the course of next year.

I'll be making early versions of the heuristics available to practitioners for use, and I'll be looking for feedback, so if you're interested please see me later or follow my blog for updates.

13. Thanks

Thanks for listening.

A link to the full project report is available on my website.

Presentation and slides will be available at http://www.digitalglam.org/crowdsourcing/talks/

For crowdsourcing research updates follow www.digitalglam.org   @donellemckinley

McKinley, D. (2013). *Why evaluation isn't a party at the end: Evaluating crowdsourcing websites* [presentation]. National Digital Forum Conference, 27 November 2013, Wellington, New Zealand. Retrieved from http://www.digitalglam.org/crowdsourcing/talks/

14. References

Bonney et al. (2009). *Public Participation in Scientific Research* (A CAISE Inquiry Group Report). Washington DC: Center for Advancement of Informal Science Education. Retrieved from http://www.birds.cornell.edu/citscitoolkit/publications/CAISE-PPSR-report-2009.pdf/view

Brabham, D. (2013). *Crowdsourcing*. Cambridge, MA: MIT Press.

Carletti, L., Giannachi, G., Price, D., & McAuley, D. (2013). Digital Humanities and Crowdsourcing: an Exploration. Presented at the Museums and the Web, Portland, Oregon. Retrieved from http://mw2013.museumsandtheweb.com/paper/digital-humanities-and-crowdsourcing-an-exploration-4/

Cockton, G., Woolrych, A., Hornbaek, K., & Frokjaer, E. (2012). Inspection-Based Evaluations. In J. A. Jacko (Ed.), *The human-computer interaction handbook : fundamentals, evolving technologies, and emerging applications* (pp. 1280–1296). Boca Raton, FL: CRC Press.

Dunn, S., & Hedges, M. (2012). *Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research*. London: Centre for e-Research, Department of Digital Humanities, King's College London. Retrieved from http://crowds.cerch.kcl.ac.uk/wp-uploads/2012/12/Crowdsourcing-connected-communities.pdf

Ellis, D., Jenkins, M., Lee, W., & Stein, R. (2008). Agile Methods for Project Management. In *Museums and the Web 2008 Proceedings*. Toronto: Archives & Museum Informatics. Retrieved from http://www.museumsandtheweb.com/mw2008/papers/jenkins/jenkins.html

Holley, R. (2010). Crowdsourcing: How and Why Should Libraries Do It? *D-Lib Magazine*, *16*(3/4). doi:10.1045/march2010-holley

Howe, J. (2008). *Crowdsourcing: why the power of the crowd is driving the future of business*. New York: Three Rivers Press.

Krug, S. (2006). *Don't Make Me Think: A common sense approach to web usability* (2nd ed.). USA: New Riders.

Lascarides, M. (2012). *Next-Gen Library Design*. Chicago: ALA Tech Source.

McKinley, D. (2013). Functionality and Usability Requirements for a Crowdsourcing Task Interface that Supports Rich Data Collection and Volunteer Participation: A Case Study: The New Zealand Reading Experience Database (Submitted to the School of Information Management, Victoria University of Wellington in partial fulfilment of the requirements for the degree of Master of Information Studies). Wellington, New Zealand: Victoria University of Wellington. Retrieved from http://researcharchive.vuw.ac.nz/handle/10063/2674

Nielsen, J., & Mack, R. L. (1994). *Usability inspection methods*. New York: Wiley.

Oomen, J., & Aroyo, L. (2011). Crowdsourcing in the cultural heritage domain: opportunities and challenges. In *C&T '11 Proceedings of the 5th International Conference on Communities and Technologies.* (pp. 138–149). doi:10.1145/2103354.2103373

McKinley, D. (2013). *Why evaluation isn't a party at the end: Evaluating crowdsourcing websites* [presentation]. National Digital Forum Conference, 27 November 2013, Wellington, New Zealand. Retrieved from http://www.digitalglam.org/crowdsourcing/talks/

Owens, T. (2013). Digital Cultural Heritage and the Crowd. *Curator: The Museum Journal*, *56*(1), 121–130. doi:10.1111/cura.12012

Petrie, H., & Power, C. (2012). What do users really care about?: a comparison of usability problems found by users and experts on highly interactive websites. In *CHI '12 Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*. (pp. 2107–2116). doi:10.1145/2207676.2208363

Rogers, Y., Preece, J., & Sharp, H. (2011). *Interaction design*. Hoboken, N.J.; Chichester: Wiley ; John Wiley [distributor].

Simula, H. (2013). The Rise and Fall of Crowdsourcing? In *2013 46th Hawaii International Conference on System Sciences (HICSS)* (pp. 2783–2791). doi:10.1109/HICSS.2013.537

Vukovic, M., Kumara, S., & Greenshpan, O. (2010). Ubiquitous crowdsourcing. In *Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing - Adjunct* (pp. 523–526). New York, NY, USA: ACM. doi:10.1145/1864431.1864504

Wiggins, A., & Crowston, K. (2010). Developing a conceptual model of virtual organisations for citizen science. *International Journal of Organisational Design and Engineering*, *1*(1/2), 148. doi:10.1504/IJODE.2010.035191