Case Notes: Turning crowdsourced information into evidence trails for collection metadata

Marcus Winter
University of Brighton
Brighton, BN2 4GJ, UK
marcus.winter@brighton.ac.uk

Phil Blume
Adaptive Technologies Ltd
Brighton, BN3 1EH, UK
phil@adaptivetechnologies.com

Susan Lambert
Arts University Bournemouth
Bournemouth, BH12 5HH, UK
susan.lambert@aub.ac.uk

Lyn Pemberton
University of Brighton
Brighton, BN2 4GJ, UK
lyn.pemberton@brighton.ac.uk

Abstract
Crowdsourcing is becoming increasingly popular in the cultural heritage sector as a way to improve and extend digital collections while at the same time engaging new audiences. A key problem, particularly in crowdsourcing efforts that ask participants to contribute complex information, is how that information can feed into the collection without the risk of compromising professional standards. This paper discusses how the problem was addressed in the 10 Most Wanted project. It presents Case Notes as a mechanism for curators to validate contributions and integrate them into an evidence trail for newly discovered facts about collection items.

Keywords
Collections; crowdsourcing; user-generated content; metadata; data quality; verification

Introduction
One of the key advantages of crowdsourcing is that it combines audience engagement with the production of useful outcomes. In the context of cultural heritage this can translate into sustainable models for maintaining and extending collections by delegating some aspects of curatorial research to members of the public.

A potential downside is that the public usually lacks the expert knowledge and skills of professional curators. While it has been suggested that crowdsourcing can lead to solutions superior in quality and quantity to professional efforts [2], there are widespread concerns among professionals about data quality. Some of these concerns are highlighted in Alexandra Eveleigh's [5] discussion of participatory archives:

"User participation initiatives in archives are haunted by a fear that a contributor might be wrong, or that descriptive data might be pulled out of archival context, and that researchers using collaboratively authored resources might somehow swallow all of this without question or substantiation." [5]

From a curator's perspective, data quality and verification are critical to avoid compromising quality standards for the collection as a whole. Introducing invalid data would not only impact on the collection's value as a research resource but also undermine the institution's authority, which is a distinguishing aspect particularly for heritage organisations [8]. Data quality is also important from the perspective of volunteers, who want to be reassured that the outputs of their efforts are useful and academically valid [4].

Measures suggested in the literature to improve data quality in crowdsourcing projects can be broadly grouped into four approaches:

1. Make the task easier: break down tasks into sub-tasks and provide higher quality materials [6]
2. Train/inform volunteers: provide learning materials [3] and best practice guidelines [7]
3. Crowdsource quality control: compare results between participants [9] or set clean-up tasks [1]
4. Professional quality control: curators as gatekeepers when integrating content into collections [5].

10 Most Wanted[1] combines several of these approaches to ensure contributions meet professional standards. It trains volunteers by providing guidance and research tips and it encourages participants to critically assess each other's findings. The main responsibility of quality control rests, however, with professional curators who screen contributions and piece together key information into an investigative narrative (case notes) evidencing newly discovered facts about an collection items. The rest of this paper gives an overview of the information flow in 10 Most Wanted, discusses various aspects of case notes and concludes with a critical review.

Case Notes

Case Notes are the product of a complex process involving the advertisement of objects and related challenges (cases) on the 10 Most Wanted website, the promotion, investigation and eventual solution of cases taking part on the project's social network channels, and the aggregation and curation of contributions into archivable and publicly accessible evidence trails for discovered facts (Figure 1).

Besides their overarching purpose to turn crowdsourced information into valid collection metadata, case notes address several other crowdsourcing related aspects in the project:

- They provide an up-to-date summary of the on-going investigation so that participants and visitors can see progress without the need to search and connect individual social media posts.

[1] 10 Most Wanted is a research project exploring complex game-based crowdsourcing for collections. See http://10most.org.uk

- They record key discoveries in the museum's own domain reducing dependency on social networks' unpredictable data storage and access practices.
- They summarise evidence in a museum context by relating information to specific questions about collection items.
- They provide a platform to credit contributors for their work and thereby help to sustain motivation.

Case notes provide a well-defined check point for curators to assess the quality of contributions and construct an evidence trail that meets professional standards. They can then be archived once a case is solved and linked to from collection metadata in order to provide a publicly accessible investigative narrative.

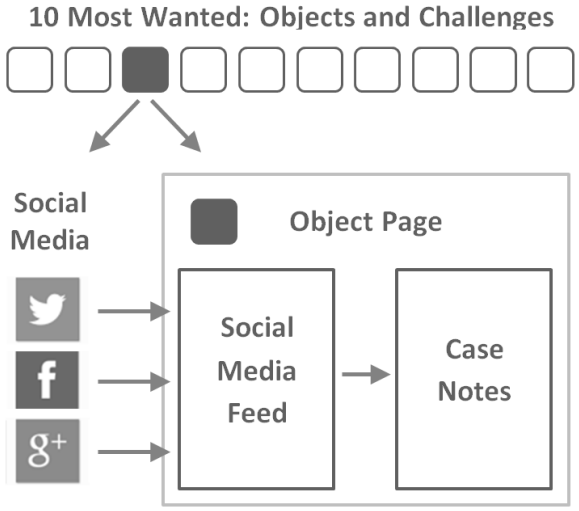

figure 1. Information flow from social media channels to curated case notes evidencing facts about collection items

Evaluation

Case notes have been used in 10 Most Wanted for over eight months to date, evidencing a wide range of newly discovered facts about collection items in a total of 15 solved cases so far. The process of maintaining case notes is well integrated into the workflow of facilitating on-going investigations on social networks and meets the requirements of curators involved in the project.

The concept was formatively evaluated in a small-scale survey involving 11 curators and other professionals working with collections. Results suggest that while most respondents agree that 10 Most Wanted is a useful approach to engaging people in new ways with collections and are comfortable with the way how it turns public contributions into formal documentation, some respondents have reservations about this aspect. While these results are not representative for the cultural heritage sector due to the small sample size, they indicate that more research is needed on the aspect of converting crowdsourced information into metadata for professionally curated collection.

Summary and conclusions

This paper discussed data quality as a key problem in crowdsourcing efforts where participants contribute complex information. It has presented case notes as a central mechanism in 10 Most Wanted to validate and integrate contributed information into evidence trails, while also addressing a range of other aspects relevant in a crowdsourcing context. Case notes are being used successfully in the 10 Most Wanted project, but there were some concerns about the concept in a small-scale formative evaluation. The results suggest that a more detailed evaluation is required to assess the validity of the concept and its acceptance among professionals.

Acknowledgements

Ten Most Wanted is supported by the Digital R&D Fund for the Arts - Nesta, Arts & Humanities Research Council and public funding by the National Lottery through Arts Council England

References

[1] Bernstein, S. (2009). Crowdsourcing the Clean-Up with Freeze Tag! Brooklyn Museum. http://www.brooklynmuseum.org/community/blogosphere/2009/05/21/crowdsourcing-the-clean-up-with-freeze-tag/

[2] Brabham, D. C. (2008). Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. Convergence: The International Journal of Research into New Media Technologies, 14(1), 75–90.

[3] Cohn, J.P. (2008). Citizen Science: Can Volunteers Do Real Research? BioScience, 58(3), 192–197.

[4] Dunn, S. and Hedges, M. (2012). Engaging the Crowd with Humanities. A scoping study. http://stuartdunn.files.wordpress.com/2013/04/crowdsourcing-connected-communities.pdf

[5] Eveleigh, A. (2012). Welcoming the World: An Exploration of Participatory Archives. Presented at Int. Council on Archives (ICA) Conference (ICA 2012).

[6] Holley, R. (2009). Many Hands Make Light Work: Public Collaborative OCR Text Correction in Australian Historic Newspapers. http://www.nla.gov.au/content/many-hands-make-light-work-public-collaborative-ocr-text-correction-in-australian-historic.

[7] Kidd, J. (2013). Visitor Generated Content (VGC) and Ethics - what we might learn from the media and journalism. Presented at iSay: State of Things.

[8] Oomen, J. and Aroyo, L. (2011). Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges. Proc. 5th International Conference on Communities and Technologies, pp. 138–149.

[9] Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K., Szalay, A. S. and Vandenberg, J. (2010). Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers. Astronomy Education Review, 9(1), 1-18.