Title
Crowd-sorting: reducing bias in decision making through consensus generated crowdsourced spatial information

Journal
International Conference on GIScience Short Paper Proceedings, 1(1)

Authors
Mcnair, Hamish
Arnold, Lesley

Publication Date
2016

DOI
10.21433/B3115wt35578

Peer reviewed

Crowd-sorting: reducing bias in decision making through consensus generated crowdsourced spatial information

H. D. McNair[1], L. M. Arnold[2]

[1]University of Canterbury, Private Bag 4800, Christchurch 8020, New Zealand
Email: hamish.mcnair@pg.canterbury.ac.nz

[2]Curtin University, Kent Street, Bentley, 6102, Western Australia
Email: lesley.arnold@curtin.edu.au

[1, 2] Cooperative Research Centre for Spatial Information, Australia

1. Introduction

Crowdsourcing complements existing planning processes. It enables people to contribute information to decision-making processes in a way that promotes outcomes better suited to their needs. People have a unique insight into the world they know, interact with, and live in, and therefore have an affinity with local issues. However, by the time traditional public consultation (e.g. town hall meeting) is undertaken, unfavourable planning decisions may have inadvertently been made. Potential oversights can be mitigated by including insights from the crowd earlier in the planning process.

Spatial data and information is used to model a truth that represents the existing landscape and identifies issues that the planning process needs to address. Spatial data and information, however, are fundamentally distinct. Data is an abstraction of the real world based on measured properties, whereas information concerns the interpretation of a scenario or event – whether by way of a first-hand account or analysis of data (Ackoff 1989). Experts use interpreted information to decide what issues should be addressed by the planning process and how. On the other hand, crowdsourced spatial information can be obtained directly from the crowd, not via expert processing. This *crowd truth* offers a balance to the *single truth* that may result from mathematical data processing (Aroya and Welty 2015).

While crowdsourced spatial information can help define the needs of a community it must be compatible with existing processes to influence change. This paper introduces Sensibel as a mechanism for crowdsourcing data and information about the current performance of cycling infrastructure. We then present an approach to *crowd-sorting* this information to increase both its usability and trustworthiness. Crowd-sorting uses the spatial characteristics of the information alongside its provenance. It considers how the crowd interacts with individual observations and the areas they describe. The aim is to use these interactions to crowd-sort contributions and produce more constructive inputs to the planning process.

1.1 Crowdsourcing in Planning

Inclusion of crowdsourcing in planning processes often overlooks local insights by focussing on rudimentary data elements, not prevailing local perspectives. Such a process is depicted in Figure 1. For example, assessing cycle routes using crowdsourced data from Strava (http://strava.com) simply shows which routes people use the most. The issue being that high volume does not necessarily equate to a well performing route. Incorporating crowdsourced information early in the planning process may better address the needs of the community. Brown (2015) indicates that land use planning can be improved by spatially arranging comments from the public. This spatial reference provides a link between data centric processes and crowdsourced information. A similar approach was applied during the formation of Helsinki's master plan (Kahlia-Tani *et al*. 2016). However, distilling public contributions into discernible courses of action is challenging given the potentially broad range of comments.

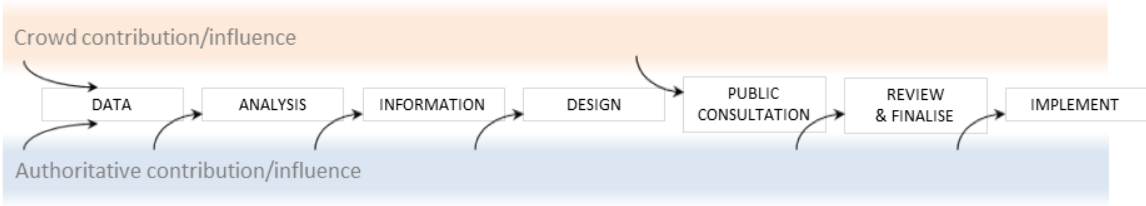

Figure 1. Contributions and influences in the planning process.

2. Generating Consensus Using Sensibel and PROV

Sensibel (https://fabriko.squarespace.com/sensibel) allows the crowd to define locations where cycle infrastructure is performing above or below expectation. The bicycle bell sized unit attaches to a cyclist's handlebars and has positive and negative buttons which, when pressed, indicate the cyclist's opinion of that location. The hardware synchs via Bluetooth with the associated mobile app. The smartphone's GPS captures the route (line feature) and the location of individual positive/negative points. Data is submitted to a central repository where spatial clustering of all cyclists' point data produces locations that the crowd can comment on via a web app. They can also agree (up vote) or disagree (down vote) with these opinions.

A possible use case is presented in Figure 2 where users **a** and **b** have negative experiences and submit points representing these (**P_a**, **P_b**). A cluster (**Cl_1**) forms at this location. Users **e** and **f** also pass through the area but do not contribute points. Users **c** and **d** have not passed through this area but may contribute opinions via web app. All contributors (**a** through **f**) are able to submit comments about the area the cluster represents and up or down vote these comments according to their opinion of them.

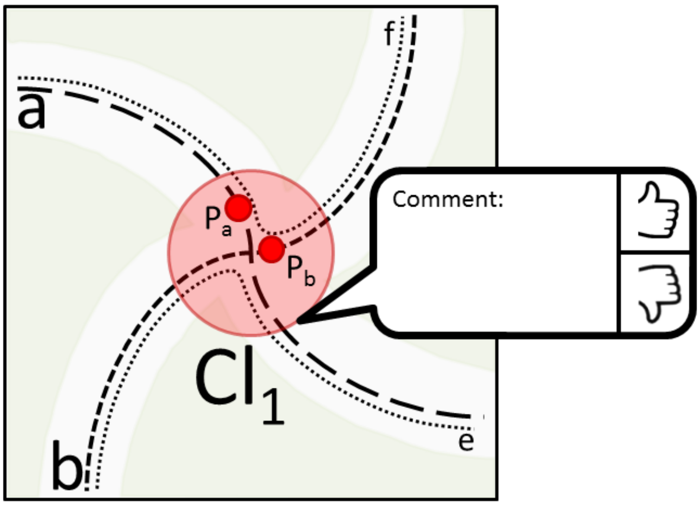

Figure 2. Sensibel in action.

The elements contributing to the formation of crowd consensus can be assessed based on how information is created and used. Provenance details an entity's creation and subsequent use or modification. The W3C developed PROV Data Model (Moreau and Missier 2013) enables detailed provenance information to be recorded and its semantic capabilities permit analysis of the relationships between descriptive elements. PROV defines three general classes – entity, activity and agent:

- An **entity** (real, digital, or otherwise). For Sensibel, entities include the data/information representing **positive/negative points**, **cyclists' routes**, **comments**, and **up/down votes**.
- An **activity** uses or acts upon entities. Users **submit** positive/negative points, **pass through** cluster areas, **contribute** comments, or **up/down vote** comments.
- An **agent** has authority over activities performed or entities used. These are Sensibel **users**.

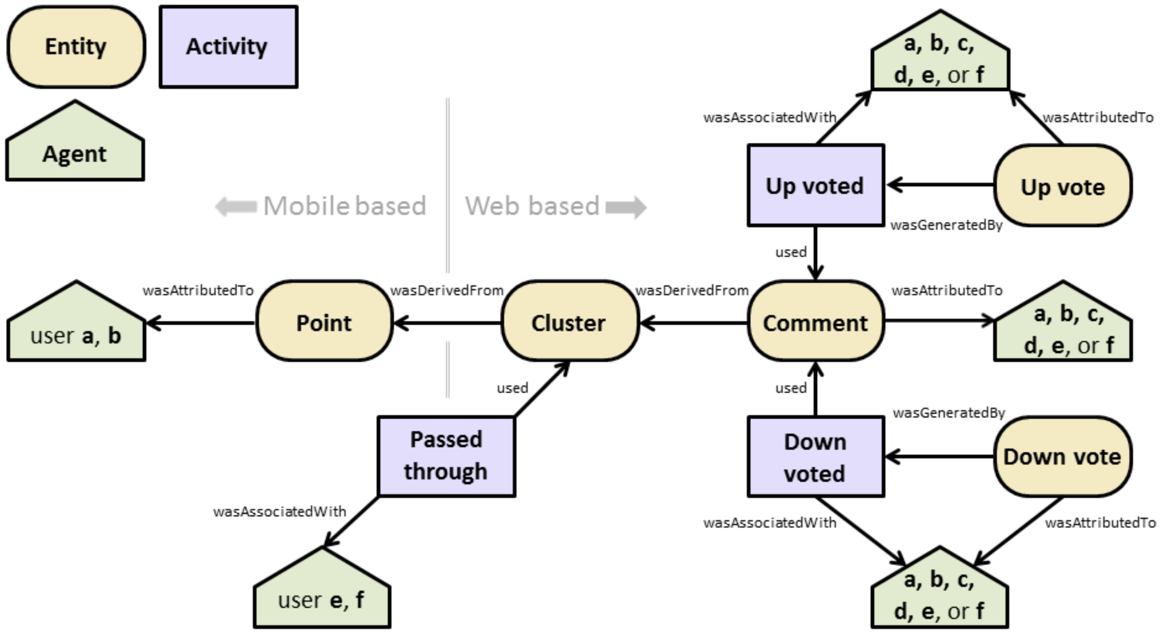

Figure 3. A simplified provenance graph detailing the Sensibel use case from Figure 2.

By defining types of user we can investigate how different kinds of interactions influence the resulting information. We propose 3 tiers of participation and their likely contribution to crowd-sorted information:

- Active-participants contribute points that form clusters (e.g. users **a** and **b**). These cyclists identify something in the area they consider important. We foresee their comments and votes as providing valuable insights into positive or negative issues associated with a site.
- Passive-participants pass through a cluster-area without identifying issues (e.g. users **e** and **f**). We predict they will have enough knowledge of the area to provide supporting evidence via web app voting.
- Remote-participants aren't registered as having visited the area (e.g. users **c** and **d**). Their comments may prove useful, but we are reliant on those familiar with the area verifying their opinions via web app voting.

PROV permits filtering by these 3 tiers – the three different types of **agent** in Figure 3 – so we can observe the impact each has on crowd consensus. This enables the contributions of each type of user to be assessed so emphasis can be placed on the most relevant contributions at each stage in the Sensibel system. It also allows for a broader examination of the importance associated with being in the place about which you are providing information. This concept was employed by Celino (2015) when using the crowd and an individual's location (with PROV) to curate a crowdsourced spatial data set. In a planning context this allows separation of the perspectives of active-participants who have recently interacted with the site and remote-participants providing comment based on their general – and potentially dated – opinion of the area.

Crowd-sorting seeks to produce inputs to the planning process that represent the views of the crowd in a format that complements existing planning processes, as illustrated by Figure 4. The top comments for each cluster are assigned its location and can be used in conjunction with other spatial data sets. For example, a spatial join with asset data sets may highlight reoccurring issues (such as an aversion to a type of railway crossing) that induces improvements in infrastructure design.

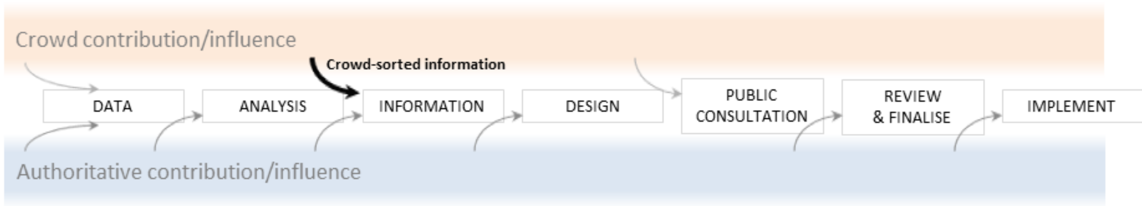

Figure 4. Crowd-sorted information in the planning process.

3. Conclusion and future work

This paper discussed the value of including crowdsourced spatial information in the planning process and introduced Sensibel as a system for collecting crowdsourced geographic data and information relating to cycle infrastructure performance. Crowd-sorting is a mechanism by which these contributions can be used to produce more representative and constructive inputs to the planning process.

The next stage is to collect data and information from a test group of users. Planners will be engaged to determine characteristics of the information that should be emphasised in order for it to have the greatest impact in the planning process. Once the elements most conducive to producing crowd consensus have been identified the semantic capabilities of PROV can be implemented to achieve effective and practical information sorting.

Acknowledgements

This work has been supported by the Cooperative Research Centre for Spatial Information (CRCSI), whose activities are funded by the Australian Government.

References

Ackoff RL, 1989. From data to wisdom. *Journal of applied systems analysis*, 16(1):3-9.

Aroyo L and Welty C, 2015. Truth is a lie: Crowd truth and the seven myths of human annotation. *AI Magazine*, 36(1):15-24.

Brown G, 2015. Engaging the wisdom of crowds and public judgement for land use planning using public participation geographic information systems. *Australian Planner*, 52(3):199-209.

Celino I, 2015. Geospatial dataset curation through a location-based game. *Semantic Web*, 6(2):121-130.

Kahila-Tani M, Broberg A, Kyttä M and Tyger T, 2015. Let the citizens map—public participation GIS as a planning support system in the Helsinki master plan process. *Planning Practice & Research*, 31(2):1-20.

Moreau L and Missier P (eds), 2013. PROV-DM: The PROV Data Model. World Wide Web Consortium.