

Human Guinea Pigs and Casual Collaborators: Crowd Sourcing Data for Archaeology

Dr Anthony Masinton

As data acquisition techniques become finer and finer grained, allowing for more detailed recording, analysis and eventual interpretation, opportunities and needs arise for dividing both the acquisition and the processing of that data into manageable portions while maintaining control of the overall project. Enter crowdsourcing. While more strictly science-based projects within high-profile organisations such as NASA and SETI have been crowdsourcing their data processing for many years, archaeology has yet to grasp its potential.

In 2010 and 2011/12 I made tentative forays into DIY crowd sourcing archaeological data with two small-scale projects. They had different reasons for turning to 'the crowd'. The first project, investigating the use of space at medieval Bodiam Castle, derived its data entirely from crowd sourcing. A student and I invited volunteers to walk through a virtual model of the castle online while we recorded their paths through the space (effectively turning them into human 'guinea pigs' negotiating a maze). We had over 600 participants which allowed us to 'ground truth' the way more conventional spatial analysis predicted the space would organise its occupants. Despite this success there were a number of unforeseen problems which provided a valuable (although difficult!) learning experience.

The second project sought crowd sourced assistance with processing a pre-existing, complex dataset which required the attention of human minds, rather than simply being processed by computer algorithm. This project involved tracing the lines of a very complex collection of medieval graffiti at Durham Cathedral. This project was much smaller-scale, in the end involving a 'crowd' of only 15, but the tools and methods developed were conceived within a crowd sourcing paradigm. Unlike the Bodiam project, the participants here were seen as collaborators or colleagues - working with me as opposed to for me. Again, the project was successful, but here too I had to manage a number of unexpected challenges.

While these two projects are interesting individually, it is insights they offer together into the practical experience of designing and managing crowd sourcing projects that make them valuable. They demonstrate that small-scale projects can and should be implemented in the humanities - that huge budgets and vast amounts of time are not always necessary. However, they also show that issues of design, aesthetics, quality control and 'customer' relations', which scholars working in the humanities may not be used to considering, are vital to the success of the project. By comparing my projects with larger, more established projects I have developed a set of working principles or guidelines which I hope will aid the effective, useful and financially-efficient use of crowd sourcing within the unique environment of the humanities.

1. The research question - its goals and the nature of its data - is of first importance. If the researcher's question is not compelling and their data not capable of division into many, easily manageable pieces, crowd sourcing is not the ideal way forward.

2. The methodology - the way in which you hope 'the crowd' will participate - is of first importance.

3. Design, polish and support matter.

4. You must reward your participants. This reward can take many forms, but the participants must feel they are gaining something of value in order for them to invest their time and resources in your work.

Crowd sourcing requires clarity of vision in your research as well as technical (and technological) mastery which engages your potential participants. There are numerous pitfalls to avoid, but if done well, it is an undeniably valuable tool.