Request for Proposals: Text-based Crowdsourcing at the University of Minnesota

As the size and complexity of digital datasets increase, dramatically new strategies in computational thinking are needed to transform data into knowledge. The Zooniverse platform for citizen science (Zooniverse.org) provides a unique crowdsourcing solution. It harnesses the visual processing capabilities of over a million volunteers worldwide who record observations on rich data sets from across many disciplines, such as the morphological classification of galaxies or the transcription of pages from old ship's logs from World War I. Aggregating the “votes” of many volunteer observations into a consensus value (crowdsourcing) has been shown to be at least as good as classifications by experts and can outperform computational algorithms asked to assess complex data. A data processing project that might take your research team years to finish can take mere months with the power of crowdsourcing. The Zooniverse currently hosts twenty-five projects, each conducted by an independent team of researchers who analyze the crowdsourced data and publish results in peer-reviewed journals.

Over the past few years, Zooniverse has mounted a series of high-profile text-based research projects that successfully use the crowdsourcing technique including Old Weather (www.oldweather.org), Notes From Nature (www.notesfromnature.org), Ancient Lives (www.ancientlives.org), Operation War Diary (www.operationwardiary.org), and What’s the Score (www.whats-the-score.org). Now, Zooniverse is expanding its digital humanities offerings under the brand “Zoomanities”. The latest project under development through this initiative is a full-text manuscript transcription project with Tate Britain, focused on artists’ sketchbooks and personal papers.

The University of Minnesota, through the Zooniverse@UMN initiative, has recently joined forces with the Zoomanities initiative to enhance and enable UMN text-based research projects. The new Zooniverse@UMN initiative also seeks to provide education opportunities with the Zooniverse projects both in the classroom and through project-based work for undergraduate and graduate students.

As part of this transdisciplinary Zooniverse@UMN initiative, we are soliciting projects for the Zoomanities effort. If you have a text-based collection that has already been digitized (e.g. scanned and in some image format) and you would like us to develop a website to harness the power of hundreds of thousands of volunteers to help transcribe your texts or annotate any aspects of your collection with metadata, please complete the Project Proposal Form below.

At any time during the process, if you have questions, please don’t hesitate to send email to zoomanities@umn.edu.

This RFP is sponsored by the Zooniverse@UMN initiative funded by the Office of the Vice President for Research, the University Libraries, the Colleges of Biological Sciences, Liberal Arts, and Science & Engineering, and the University of Minnesota Informatics Institute.
Project Selection Criteria

Essentials:

- Quality of research question, e.g. output must be publication-worthy
- Compelling and accessible research questions within and beyond the academy
- At least one of the proposed project leads must be a Faculty or Staff member affiliated with the University of Minnesota through the duration of the proposed project
- In keeping with the open data spirit of the Citizen Science Alliance (the parent organization of Zooniverse), all data resulting from the crowdsourcing process must eventually be made public (based on the standards and conventions of open data in the academic field of the proposed project)
- A significant portion of the material to be analyzed must already be digitized
- Technical feasibility

Desirables:

- Projects that have interdisciplinary potential or applications
- We particularly welcome project partners who would wish to explore embedding their project in undergraduate curricula or who propose other education components

Timeline

Proposals will be due between November 24th and December 15th 2014. For evaluation purposes, early submission is strongly encouraged. Please submit responses to the questions below as well as related material (see questions 12 and 13 below) as a single PDF to zoomanities@umn.edu. Proposals will be reviewed by both the Zooniverse@UMN Project Team and independent experts. Successful applicants will be notified by February 1st 2015.

For organizational purposes an Expression of Interest (EoI) emailed by November 3rd 2014 to zoomanities@umn.edu is strongly encouraged. The EoI must include the UMN contact name and a short abstract (150-200 words) of the proposed project.

Zooniverse@UMN Project Team/Advisory Board

- Lucy Fortson, (Associate Head and Associate Professor, School of Physics and Astronomy)
- Claudia Neuhauser (Director of UMN Informatics Institute)
- Mark Sanders (Director of UMN Imaging Centers)
- Joy Kirchner (Associate University Librarian, University Libraries)
- Nita Krevans (Associate Professor and Chair, Classical and Near Eastern Studies)
- Craig Packer (Professor, Department of Ecology, Evolution and Behavior)
- Loren Terveen (Professor, Department of Computer Science and Engineering)
- Justin Schell (Digital Arts Sciences + Humanities, University Libraries)
- Carolyn Bishoff (Physics, Astronomy and Earth Sciences Librarian)
- Victoria Van Hyning (Oxford University - External Advisor)
Project Proposal

I. General Information

Project Title:

Abstract:

Team Information (include contact information of at least one UMN faculty or staff member who will be the Project Lead):

Relevant Websites:

II. Project Specifics

1. Provide a description of the humanities question or subject area to be addressed by your text-based project. What data will you gather through crowdsourcing and how will you use it? Why is crowdsourcing the appropriate or best way to gather your data? (no more than 500 words)

2. Describe the tasks you envisage volunteer participants carrying out - including (a) any required specialist knowledge, (b) minimum requirements for success, (c) desired outcomes. If you or your colleagues have carried out the proposed task yourself, how much time do you estimate each individual task will take? What would a minimally successful project look like? (300 words)

3. What automatic processing routines exist which attempt to solve the problem being addressed (e.g. OCR)? Explain what elements of your data cannot be fully processed by these methods or why your project cannot use these methods. (150 words)

4. Describe the nature of the scanned material that would be used in the proposed project - include (a) its format (file type, size, number of files) (b) its likely copyright owners (i.e.: public domain; researcher's own work; owned by X) and if not public domain, why it can be used for this project (c) its availability (is it archived data or still in the process of being collected/digitized?). Recall: For your project to be considered, a significant amount of the data needs to be already digitized. (150 words)

5. Who will carry out the analysis of the data collected by the proposed project, particularly during any proprietary period? Zooniverse, and in particular Zooniverse@UMN, intends to work with their project partners to aggregate and interpret data, but it is very useful for the proposing team to have some in-house data manipulation capabilities. (150 words)
6. All data from Zooniverse projects resulting from the crowdsourcing process must eventually be made public. What final format (catalogue, annotated images, query tool) would be needed? What are the anticipated final outcomes (e.g., papers, catalogues)? Are the results of this project likely to be of interest to researchers beyond your own field? If so, how? (300 words)

7. Are there potential extensions to the project that you have in mind? For instance, running the proposed task on a different set of materials or running a different set of tasks on the same materials. (150 words)

8. The primary goal of our projects is to produce academic research, but projects can have significant educational impact as well. Engaging the community is an excellent way of ensuring they remain committed to gathering data for you. Are there members of your team willing to write blog posts, join forum discussions and otherwise take part in outreach? Does the project tie in with any public engagement or education activities you are already involved with? (300 words)

9. Are there any clear connections that would help with marketing your project and recruiting volunteers? Existing interest groups, local contacts, or media links? (150 words)

10. Does your project look like any of the current Zooniverse projects? If so, which one(s)?

11. Does your project already have existing funding? If so, please summarize what aspects are currently funded. (100 words)

12. Attach a bibliography of papers and conference proceedings relevant to the proposed project.

13. Attach at least one image of the type of data/image that people will see when they work on your project.

Submit your responses to the questions as well as material from questions 12 & 13 as a single PDF to zoomanities@umn.edu

If you have any questions regarding the development of your proposal or how to respond to any of the above questions, please email zoomanities@umn.edu.