

Citizen Archaeologists. Online Collaborative Research about the Human Past

ANDREW BEVAN, University College London

DANIEL PETT, British Museum

CHIARA BONACCHI, University College London

ADI KEINAN-SCHOONBAERT, University College London

DANIEL LOMBRAÑA GONZÁLEZ, crowdcrafting.org

RACHAEL SPARKS, University College London

JENNIFER WEXLER, British Museum

NEIL WILKIN, British Museum

ABSTRACT

Archaeology has a long tradition of volunteer involvement but also faces considerable challenges in protecting and understanding a geographically widespread, rapidly dwindling and ever threatened cultural resource. This paper considers a newly launched, multi-application crowdsourcing project called MicroPasts that enables both community-led and massive online contributions to high quality research in archaeology, history and heritage. We reflect on preliminary results from this initiative with a focus on the technical challenges, quality control issues and contributors motivations.

1. INTRODUCTION. VOLUNTEER RESEARCH INTO HUMAN HISTORY

Archaeology has long been successful in piquing the interest of full-time practitioners, organised volunteer societies and the wider public alike. This enthusiasm is especially clear in the United Kingdom, where the subject has benefited from an enduring tradition of volunteer fieldwork, special interest groups and dedicated media coverage, in step with similarly strong and long-established UK

citizen involvement in other fields such as environmental, meteorological and astronomical monitoring (Roy et al. 2012). Archaeologists (in common with other fields such as biodiversity studies) seek to protect and understand a massive, geographically-scattered and constantly-threatened resource, with what have typically been only small amounts of public or private money. This paper introduces and discusses early results from a project called MicroPasts (micropasts.org) which seeks to bring together full-time archaeologists, historians, heritage specialists, volunteer archaeological societies and other interested members of the public to collaborate in both old and new forms of research about human history worldwide. Results from the first phases of this project now allow us to reflect critically on technical issues associated with delivering a complex project of this kind, the pros and cons of different quality control strategies and the kinds of contributor support that we have attracted so far.

There are compelling reasons to distribute responsibility for archaeological and document-based research beyond a rarified group of traditional specialists. In the last few years, the opportunities provided by digital technologies for wider public engagement with archaeology have attracted considerable attention (e.g. Bonacchi 2012) and, alongside these developments, there has also been an increasing focus on more reproducible forms of archaeological practice, as well as more open, participatory forms of data creation (Kansa et al. 2011; Lake 2012). All of this is in-step with much wider shifts in the sciences, social sciences and humanities. Early examples of online crowd-sourcing in archaeology and related subjects have focused on locating and photographing prehistoric monuments (the Megalithic Portal: www.megalithic.co.uk), identifying archaeological features on satellite imagery (Field Expedition Mongolia: exploration.nationalgeographic.com), pooling wartime tangible heritage (the Great War Archive: www.oucs.ox.ac.uk/ww1lit/gwa), deciphering papyri (the Ancient Lives Project: ancientlives.org), interrogating built architecture (heritagetogether.org) engagement with indigenous intellectual property (Mukurto: mukurto.org), transcribing old excavations records (the Ur excavations: urcrowdsource.org), mapping and disambiguating ancient place-names (Pleiades: pleiades.stoa.org) and recording metal artefacts (the Portable Antiquities Scheme: finds.org.uk). What is striking however is that most efforts so far have involved (a) bespoke, single-purpose crowd-sourcing efforts rather than multi-application platforms that might foster cross-over interest among archaeological enthusiasts, and (b) largely one-way models of participation (Simon 2010: 187; Dunn and Hedges 2012). Beyond this, we believe it will be enormously beneficial to provide opportunities for people traditionally distinguished as ‘academic archaeologists’, fieldwork ‘professionals’ and ‘amateurs’ not only to collaboratively produce research data across a wide variety of applications, but also to develop new research initiatives collectively, and resource them via crowd-funding appeals.

MicroPasts is a web platform that we have been developing with this large set of archaeological goals in mind. The project began in October 2013 and was in a development and testing phase until mid-April 2014 when its first crowd-sourcing applications were launched (figure 1). MicroPasts includes several distinct components, built on different technical infrastructures, but with the goal of creating a coherent set of linked online initiatives. The open source PyBossa framework (pybossa.com) is used for handling MicroPasts’ task scheduling and task presentation challenges, whilst the community forum uses Discourse (github.com/discourse/discourse), the research blog

section is supported by WordPress (wordpress.org), and the crowd-funding component has forked a version of Neighbor.ly (github.com/neighborly/neighborly, see also github.com/catarse/catarse). Users can contribute either anonymously or publicly, but for those contributors who wish to declare their identities in some fashion, we work with a combination of social logins (Facebook, Twitter, Google) and user avatars (Gravatar). All the project's datasets are made publicly available under Creative Commons licenses (CC0 or CC-BY in most cases) while the software enabling the site is also open-licensed and publicly available (github.com/MicroPasts). Hence, while the focus for this paper is primarily on crowd-sourced data collection, the broader rationale of MicroPasts is to connect three hitherto largely distinct domains -- traditional crowd-sourced science (mainly researcher-led), collaborative project design (involving both traditional academics and other interested community groups) and crowd-funding appeals (for the aforementioned collaborations) -- in ways that should be mutually supportive of one another.

One of the reasons that we expect these three different aspects to be greater than the sum of their parts is that they should involve interesting overlaps and distinctions with regard to contributors who participate in them. Here we have in mind Haythornthwaite (2009) and others' emphasis on the social continuum from 'crowd' (largely anonymous and fleetingly involved) to 'community' (identifiable avatars, repeatedly involved, with clearer group consciousness), and hope that a more complicated mix of online crowd-sourcing and community micro-funding applications will encourage people to move to and fro along this continuum of personal involvement as they wish. We also can draw on a range of pre-existing communities of archaeological and historical interest, such as organised volunteer societies. This having been said, there are also reasons to be wary. Right now, 'crowd', 'community' and 'citizen' happen to be trendy words to conjure with, in many different sectors of society, and MicroPasts is not the only web science initiative to name-check all three. However, like all connotative vocabularies, the use-value of these terms will slowly decline, as people become first familiar with them, then bored, and as examples of more cynical use proliferate. Unfortunately, examples of 'citizen sheen' or 'community-washing' are likely to become common (and for 'crowdwashing' in business, see Hück 2011). Despite the techno-utopian views that sometimes proliferate with respect to web-enabled collaboration and network societies, there are some basic tensions here that recapitulate those we find in offline interpersonal relationships (Bevan 2012b). For example, micro-funding elicits mixed feelings, as this form of entrepreneurship can sometimes feel poorly aligned with purist goals of public cooperation (see also Wheat et al. 2013). Other concerns that are likely to surface (and indeed already do informally on discussion forums) include the worry that crowd-sourcing will 'de-professionalise' archaeology or that crowd-sourced 'big data' is likely to encourage a very unreflective kind of empiricism in which information about the past is increasingly free and constantly gathered, but largely assumed to speak for itself (see Johnson 2011, also Bevan in press).

The screenshot shows the MicroPasts website interface. At the top, there is a navigation bar with links for 'MicroPasts home', 'Contributors', 'Applications', 'About', 'Forum', and 'Sign in'. The main heading is 'MicroPasts: Crowd-sourcing'. Below this, a section titled 'Help us gather high quality research data about our human history' explains the project's goals and lists benefits for contributors, such as having a direct impact on research and developing skills. A 'Start' button is provided for this section. Below the introductory text, four featured applications are displayed, each with a representative image and a 'Start' button:

- Photo Masking of British Museum Bronze Age Objects:** Masking photos of British Museum Bronze Age objects to support creation of 3D models.
- British Museum Bronze Age Index Drawer B8:** British Museum Bronze Age Card Index transcription for drawer B8.
- British Museum Bronze Age Index Drawer B16:** British Museum Bronze Age Card Index transcription for drawer B16.
- British Museum Bronze Age Index Drawer A9:** British Museum Bronze Age Card Index transcription for drawer A9.

At the bottom of the screenshot, a section titled 'Most Active Volunteers' displays a row of profile pictures of contributors.

Figure 1. The MicroPasts crowd-sourcing site with several different applications to choose from and a showcase of the most active contributors (crowdsourced.micropasts.org).

2. METHODS

Although MicroPasts' ultimate goal is to enable archaeological research regardless of country of origin, our first crowdsourcing applications have concentrated on the British Bronze Age as this provides a widely-popular period for encouraging participation (beginning with but not limited to interest in the UK) and a transferable pilot study. These first applications have included three main volunteer components: (a) the transcription of a hard copy catalogue of archaeological finds from Bronze Age Britain, (b) georeferencing of the same catalogued objects when evidence of the archaeological findspot is known, and (c) the construction of 3D models of such objects via careful masking of digital photographs.

The first two components have been implemented in a single crowd-sourcing application and aim to digitise a national record of some 30,000 Bronze Age metal finds (covering ca. 2500-800 BC). Metal finds offer one of the sharpest kinds of dating evidence (Needham et al. 1998) of any artefact types from this phase of British prehistory and, in addition, are crucial for reconstructing Bronze Age social, economic, technological, political and ritual life. While all prehistoric bronze finds in

England since 2003 have been recorded as part of the Portable Antiquities Scheme (PAS, finds.org.uk, see also Bland 2005), the 30,000 or so found during the 19th and 20th centuries languish in this hard copy catalogue which contains information on each artefact's findspot, type, condition and current whereabouts, alongside detailed line drawings and further information on the context of discovery. The index itself was a major archaeological initiative begun in 1913 and then moved to the British Museum in the 1920s, where it was maintained as a key national heritage inventory. Combining this catalogue with the PAS data will produce a near-comprehensive record of English Bronze Age metalwork for the first time and, to our knowledge, constitute the densest georeferenced database of archaeological metal artefacts worldwide. These finds also enable large-scale spatial comparison of findspot distributions and typologies (e.g. Bevan 2012a). It is also worth noting that the Portable Antiquities Scheme itself has been a wildly successful experiment in public participation for over 15 years (and occasionally contentious given its engagement with metal-detectors who are an often vilified community allowed to operate legally in the UK, see Bland 2005), and one that has also moved in recent years to an online contributive model.

The back and front of these index cards have been digitised using a fast sheet-feed scanner and then combined via Python scripting into one image that is uploaded to Flickr. The newly created sets of images are then served to the crowd-sourcing application in batches that correspond to the drawers in the original filing cabinet. Figure 2 shows an example of a transcription task. Such transcription is not something a computer can achieve easily: the cards are hand-written in various writing styles, there are notes made in odd places on the card, some include hand-drawn sketches, there is a slight variation in card layout, etc. While there are already several successful crowd-sourced transcription projects in the humanities (e.g. Brohan et al. 2009; Causer et al. 2012), the British Museum transcription tasks are challenging because of their mix of both structured (artefact lengths, widths etc. recorded in consistent locations on the card) and unstructured information (marginalia). An additional aim is to convert the locations of the finds recorded in words on the cards into geographic coordinates with an associated rough estimate of positional error. This is something that could be batch geocoded offline, but we think that there are advantages to having several human contributors implement placename look-up and then, if they feel it is appropriate, modify this further via the resulting drag-and-drop marker on a world map (OpenLayers3, Nominatim).

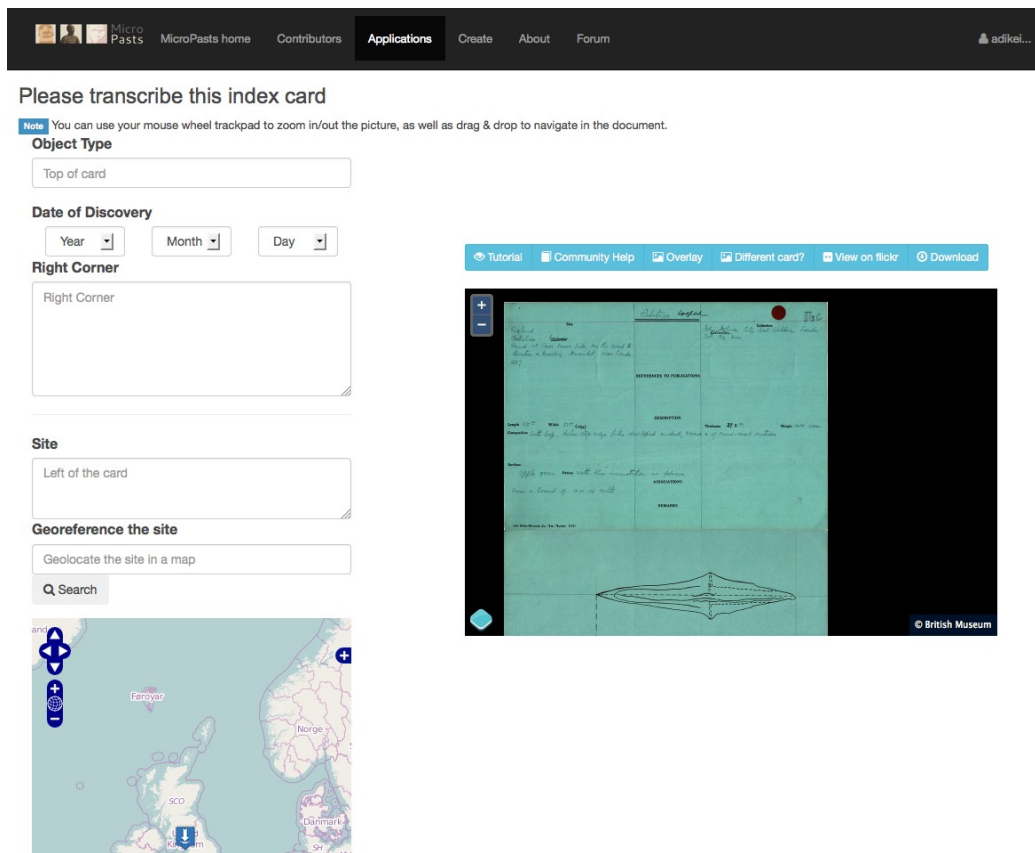


Figure 2. Crowd-sourced transcription of index cards (with georeferencing bottom left).

A further goal of MicroPasts crowd-sourcing efforts is to create a large series of research-quality 3D models of artefacts held in museum collections. For several years, there have rightly been calls for participatory object digitisation in the museum sector (Terras 2010) and it has equally been clear that structure-from-motion/multi-view stereo methods (hereafter SfM) were one attractive way to capture such models (Snavely et al. 2008). SfM is a computer vision technique that involves the creation of 3D colour-realistic models from ordinary digital photographs, often taken in ordinary conditions with ordinary cameras (for early archaeological applications, see Ducke et al. 2011; Verhoeven et al. 2012). It offers a good complement to other 3D modelling approaches (James and Robson 2012), but also has its own unique selling points: unlike traditional photogrammetry, little or no prior control of camera position is necessary, and unlike 3D laser-scanning, no major equipment costs or setup are involved. Colour information is also co-registered as part of the model-building process, rather than draped on in a potentially inaccurate second stage. The image processing demands of SfM for object models are now met by a desktop computer running for a matter of minutes, followed by a small amount of model clean-up. More importantly, the photographs necessary for SfM can be taken by anyone with a good camera and modest prior training about the preferred number and overlap of photos. The enormous value of these models lies not only in the widespread opportunities they offer for re-use in multimedia applications and immersive 3D environments, but also because, in the large numbers potentially enabled by crowd-sourcing, they

can support considerably enhanced typological analysis via 3D point cloud geomorphometrics (e.g. MacLeod 2010; Bevan et al. 2014).

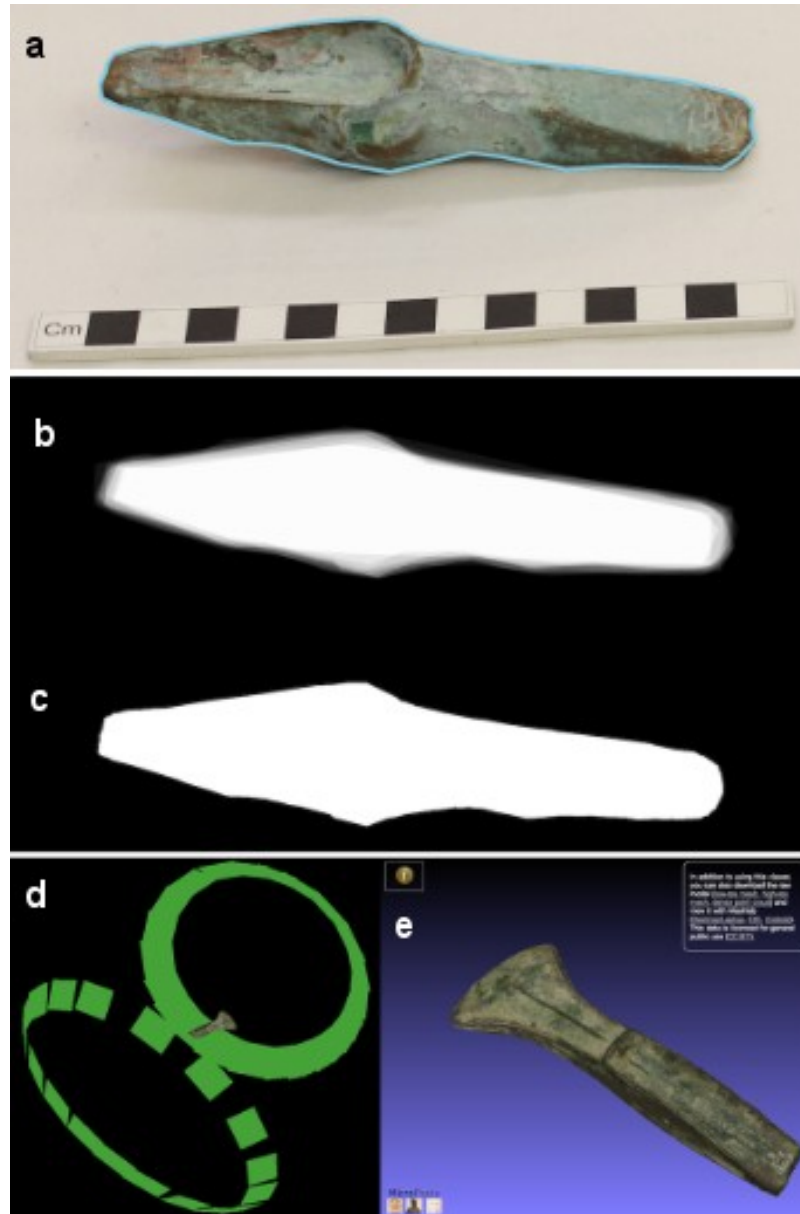


Figure 3. Crowd-sourced 3D object models: (a) an ordinary photograph taken in the British Museum of a Bronze Age axe, with a polygon outline drawn around it on the crowd-sourcing site, (b) the probabilistic raster mask from five different contributors' polygons, (c) a final binary raster mask, (d) reconstructed camera positions of multiple, masked photographs and a dense 3D point cloud of a similar object, and (e) a completed, photo-textured mesh of the object displayed to the public via a WebGL viewer, with accompanying downloadable files (NB a-c and d-e show different example axe-heads).

3D modelling via SfM will become increasingly automated, and there are already good online provisions for the public to build their own SfM models merely by uploading raw images (e.g. photosynth.net; www.123dapp.com/catch). However, better, more reproducible results can usually still be achieved offline. In particular, certain SfM approaches (e.g. PhotoScan) can exclude particular parts of each image from the feature-matching process that reconstructs original camera positions and can also mask out features from subsequent 3D model-building steps. Whilst this photo-masking step is laborious, it nonetheless produces better models than simply asking a computer to distinguish crisply between object and background on its own (at least in the present state of the art). This is especially true where the object has been flipped over onto its other side half-way through the photography session or where it has been photographed on a turntable so that, again, the background remains static while the object is seen to move. It is also worth stressing that the raw image sets are the key long-term digital resource rather than the final models, because we can anticipate that model-building algorithms will rapidly improve and we will surely want to construct fresh models in the future. Figure 3 shows an example of the role of masking in a complete SfM workflow, where contributors conduct the masking step on the MicroPasts platform by carefully drawing the objects' outline using the provided drawing and editing tools. A more general reason to familiarise volunteers with 3D modelling via photo-masking is so that, while currently we are providing images captured by museum personnel and researchers that online crowd-sourcing contributors then mask, it is highly desirable that public contributors will also be able to contribute the photographs necessary for the modelling themselves. For instance, although it may sometimes be difficult for members of the public to get the necessary access to objects in museums for good photographic capture, in contrast public visits to registered archaeological sites and landscapes is much easier and the MicroPasts project will also be looking to move in this direction of user-contributed photographs in the future (for two good case studies, see heritagetogether.org or accordproject.wordpress.com).

3. EARLY RESULTS

At the time of writing and some eight months after launch, the project has attracted well over a thousand unique contributors (800 having registered, the rest anonymous) who have completed 28 distinct applications and some 37,448 individual task runs. It is also worth bearing in mind that these tasks typically involve complex photo-masking and transcription operations rather than quick inputs, with the former taking perhaps about two minutes per task and the latter perhaps five (we are currently working on recording a clearer measure of task duration). Despite the British subject matter of our first applications, the geographical range of interest is reassuringly wide, with some two thirds of page visits coming from beyond the UK and spanning 137 different countries (figure 4). Below, we briefly raise some of the technical challenges, quality control trade-offs and aspects of contributor participation that we have identified so far.

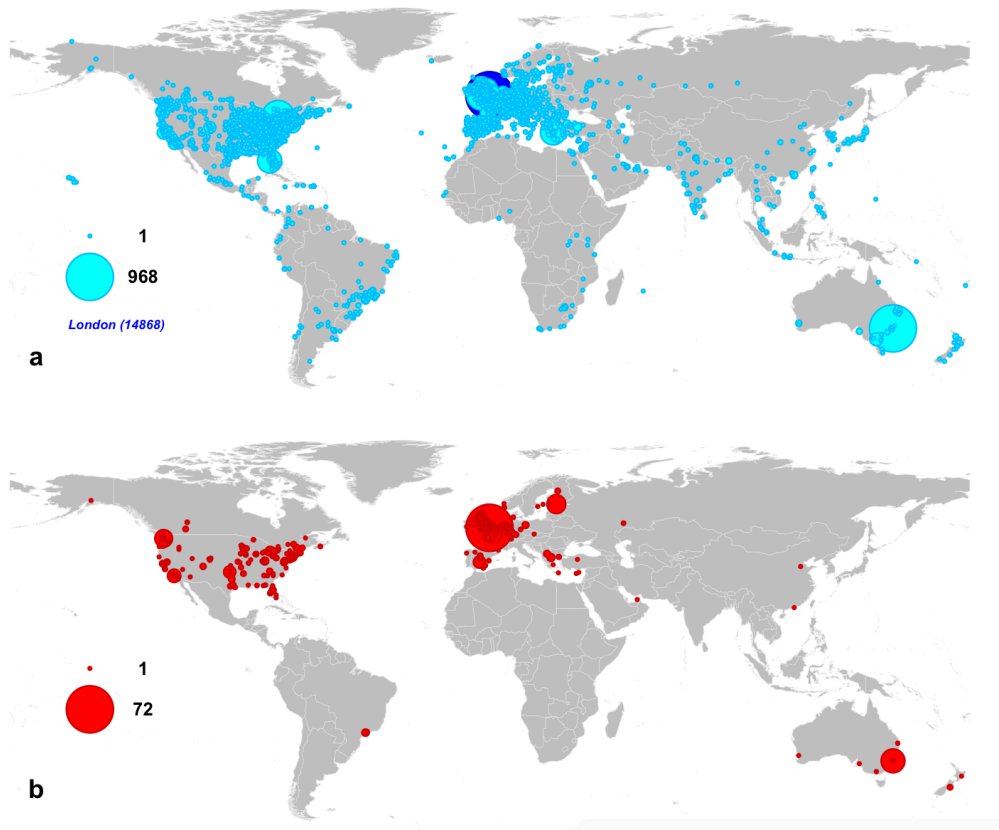


Figure 4. *Two views of the geographical reach of MicroPasts' early crowd-sourcing applications as visible via resolvable IP addresses: (a) the numbers of sessions recorded by Google Analytics between April 15 and December 10, 2014, and (b) task runs completed by anonymous users over the same time period (these represent 3.6% of total task runs). Note that while the IP addresses of registered users have not been recorded, our impression is that the geographical distribution of such contributions is similar to the above.*

3.1 Quality Control

Quality control is obviously a major challenge for any crowd-sourcing application, and it is fair to say the online transcription, georeferencing and photo-masking applications considered here involve complex rather than simple tasks, with lots of room for error and prone to considerable variation in contributor expertise. A new user is therefore provided with a tutorial of how to go about the necessary tasks and then there is always the option for them to return to the tutorial at any stage or ask questions on the community forum. Beyond this, our initial forays were designed to be conservative, to create redundancy of information and to be modifiable at several later stages (in the discussion that follows we adopt some of the terminology proposed in Allahbakhsh et al. 2013).

For photo-masking, we began by asking five different people to draw their own mask for each object photograph and then used a script to read all the resulting JSON records of contributed photo-masking tasks, extract each polygon and convert it into a black-and-white binary raster the

same size as the original photo. This was repeated for each contributor and an average taken of the five raster results resulting in a probabilistic raster mask in which pixel values near 1 implied that they were masked by everyone and those near zero that they were not (figure 3b). An arbitrary cut-off was then used (e.g. overlap in three out of five contributed polygons) to convert this to a final binary mask that can be used as input for 3D modelling (figure 3c). This proved to be a very effective method with little or no need for further post-processing, but given that the quality of user contributions has been consistently high, we have since then reduced the required number of contributions to only 2, with the second kept as back-up in case the first is inadequate.

Turning to the transcription application, we were aware from the outset that our choice of quality control mechanism for the first three drawers of the Bronze Age card catalogue was relying heavily on ‘expert review’ and would leave British Museum staff with considerable work left to do (for another transcription project where such post facto expert review is key, see Causer et al. 2012). Initially, each card was transcribed at least three times which of course also risked discouraging contributors if they feel that this is inefficient use of their time. However, for all our more recent transcription applications, we have adopted the same strategy as the for the photo-masks and only sought two contributed transcriptions per card, asking regular contributors to then help us with the consolidation task offline. We are also exploring an alternative quality control strategy in which the first contributor fills in an index card as before, but the second inputter and all others are then presented with a look-up table of previous inputs, from which they can vote for an entry or provide a wholly new one (for a similar goal achieved via online review at the Smithsonian, see transcription.si.edu). We would then lock the entry when, for example, one input option for every field has at least 3 votes (perhaps termed ‘converging agreement’). This might mean that more problematic cards could remain in circulation on the site for longer, but if discrepancies still remained for these after a certain number of tries, they could be passed automatically on to museum staff for final arbitration. In the future and in step with some recent features developed for the Pybossa framework, we will also be exploring this voting approach as one of several ways by which contributors might build online reputations, thereby enabling a further form of quality control.

3.2 Technical Challenges

Due to its diversity of purpose, MicroPasts presents significant challenges in terms of software development, especially with regard to its mixture of web technologies. We have also been keen to prioritise technical knowledge transfer within the project, making initial development of the platform a steeper learning curve than it would otherwise be, but with a firm view of the need for longer-term capacity building within archaeology. For example, we only allocated 5.5% of project budget for external software development (e.g. compared with 11.4% for the Transcribe Bentham project: Causer et al 2012: fig.1) and have only used a fraction of that so far (for the key work provided by Lombrana González on PyBossa), preferring to develop in-house amongst the core project team (who have computational backgrounds in many instances, but were not skilled in all of the necessary languages and methods beforehand) and holding the rest in reserve for troubleshooting. All project code is contributed to and updated via a GitHub repository (github.com/micropasts) in a variety of programming languages. This has enabled a truly

collaborative and open approach to application development for archaeological research. The choice of software has also raised some fundamental questions: do we create new software from scratch or modify existing software solutions for our needs? Do we seek to harness the talents of the developers who originally developed certain software or do we fork the latter and focus on in-house development? What would be the best way of storing and delivering huge numbers of images and generated 3D models? How do we ensure sustainability for our project platforms beyond an initial period of 18 months full funding?

In particular, two interesting problems are raised by the question of long-term archiving and open access to spatial data. Although we have a prior agreement to archive our datasets with the UK Archaeology Data Service (ADS, archaeologydataservice.ac.uk), it is clear that the sheer number of raw images that underpin our archival transcription and 3D modelling will make it impossible to deposit all of these with the ADS under current costing arrangements, and we are likely to be forced to archive only the final 3D models and csv files of the transcribed archival data, whilst keeping the raw images in an AWS account for the foreseeable future. This is especially unfortunate for the raw images used for 3D modelling as one might easily anticipate that SfM algorithms will improve considerable over the next few years so it is really the raw images rather than the final models that are in more urgent need of curation. A different issue is raised by our georeferenced data, with many practitioners questioning the advisability of making such information available to members of the public, given risk of known findspots encouraging renewed looting (for discussion, see Bevan 2012: 7-8). Our current applications involve only place-name geocoding and it therefore is likely to be very rare that the resulting spatial coordinates will be of sufficient accuracy to justify such concerns. At present, our policy is to enable untrammelled access to the raw crowd-sourced data, but place subsequent consolidated datasets (after expert review) under the protocols used by the PAS, in which >1km precision coordinate are provided online, but researchers need to register and request finer spatial data if they desire it.

3.3 Contributor Participation

What motivates volunteers in archaeology and history? How can crowd-sourcing support existing cultural interests and foster new ones? What do contributors and partnering institutions get from this kind of participation? There is a rapidly growing literature exploring public perceptions and experience of archaeology (both on- and offline, see Bonacchi 2012, 2014) to which we would like to add, whilst remaining as light-touch in our direct questioning of contributors as we can. The very short set of survey questions we asked after a contributor had completed their first task suggests that 74% do not work with history or archaeology in a professional day-to-day capacity (although they may well have prior experience, employment or education linking them to these subjects). The majority have also not reached the site via direct links from the research institutions involved in the project (University College London, the British Museum or the Portable Antiquities Scheme), but rather via online newspapers, magazines and other networks. It is therefore reassuring, given our initial intentions, that volunteers include many who are from outside narrow academic environments, but it is worth noting that, in contrast and partly contrary to our original expectations, organised groups conducting archaeological and historical research offline in the UK are as yet

heavily under-represented amongst active contributors. In fact, only 3% of the latter had heard of MicroPasts via organised archaeological or historical societies, despite our targeted communications. Clearly, this is something worth tracking over the longer term, but so far we are not convinced that this initial trend is solely the result of a mismatch between society members' digital skills and the kinds of online volunteering that MicroPasts is proposing (and hence not necessarily linked to socio-demographics such as age). Regardless, the very preliminary data available so far suggests a geographically dispersed and socially varied crowd of contributors.

Photo-masking appears to be the application that contributors first try on the site and the one that elicits the most inputs from one-off contributors (e.g. those who have done only 1-3 tasks). In contrast, many of the more involved contributors (e.g. those who have already done over 100 tasks) have often chosen to focus on transcription, even though, per task, this is more time-consuming and intellectually onerous. In fact, our initial impression is that the greater challenge of deciphering handwriting and complicated (often antiquated) terminology and the serendipitous discoveries that can be made over the course of transcribing multiple cards (e.g. unusual artefacts recorded on an index card, particularly skilled artefact line drawings, interesting asides such as that an artefact had been donated by Queen Victoria) are all factors behind the greater popularity of this application amongst major contributors. Furthermore (and perhaps unsurprisingly), the more intensive the task, the more contributors have sought recognition for their work, as showed by the fact that the ratio of authenticated to anonymous contributors is 2.3 for transcriptions whilst it is only 1.3 for photo-masking.

4. CONCLUSIONS

Most existing citizen science projects tend to be either single-application and single-subject or multi-application (different kinds of tasks) and domain-agnostic (spanning multiple, potentially unrelated, subject areas), but what the above discussion should emphasise is that there may well be advantages in developing more niched crowd-sourcing initiatives that are, by contrast, multi-application but domain-specific (i.e. focused on the related themes of archaeology, history and heritage). We have already experimented with the tagging of archival archaeological photographs and further applications might foster the creation of amphora 3D modelling via 2D line drawings, searches for archaeological information in online newspaper archives or on-site volunteer mapping of Medieval standing building evidence (to name just a few opportunities). As citizen science projects beyond archaeology have already made abundantly clear, it is worth thinking about modular, transferable crowd-sourcing applications, as there are increasing returns on technical investment to be had by re-tasking general applications for online transcription, image upload, photo-editing, georeferencing, etc. to new data collection priorities, and because contributors thereby also become very familiar with the different skills they require. For archaeology in particular (although the same may also be true in subjects such as taxonomic biology or palaeontology), we would further argue that bringing together archival transcription and 3D modelling (of the objects or sites referred to in those archives) is a particularly effective recipe for good research, as it creates: (a) newly quantifiable and georeferenced information from long-dormant archaeological inventories (without new fieldwork which is expensive, often

inadequately published and always destructive of the resource it explores) and (b) sample sizes of 3D models that can be analytically useful rather than just aesthetically interesting. Indeed, since archives and finds from the same archaeological site are notoriously widely distributed across different research institutions, this kind of crowd-sourced approach offers an effective way to reassemble them, in step with the arguments by Latour (1990) and others about the unusual mustering roles of certain scientific technologies. Furthermore, while there will always be a role for crowd-sourcing projects that are wholly designed by researchers in universities and museums, we also hope that by emphasising particular themes (e.g. British Bronze Age metal finds), we can build more coherent public knowledge than would be the case with more isolated data collection exercises and thereby also foster new research projects that are more collaboratively designed.

ACKNOWLEDGEMENTS

MicroPasts involves collaboration between researchers at the Institute of Archaeology, University College London and the British Museum, as well as the work of contributors worldwide. It has kindly been given an initial round of funding by the UK Arts and Humanities Research Council. We would like to thank Roger Bland, Ian Carroll, Tim Causer, Nathalie Cohen, Stuart Dunn, Susie Green, Lorna Richardson, Mia Ridge, Stuart Robson, Peter Schauer, Melissa Terras, Lisa Westcott Wilkins and Brendon Wilkins, either for comment on this draft or for other useful guidance. Much of the data discussed above was produced by both registered and anonymous contributors to the MicroPasts site, and we are very grateful for their individual help.

REFERENCES

- Allahbakhsh, M., Benatallah, B., Ignjatovic, A., Motahari-Nezhad, H.R., Bertino, E. and S. Dustdar (2013). Quality control in crowdsourcing systems. Issues and directions, *Internet Computing IEEE* 17.2: 76-81. <http://dx.doi.org/10.1109/MIC.2013.20>
- Bevan, A. (2012a). Spatial methods for analysing large-scale artefact inventories, *Antiquity* 86.332: 492-506.
- Bevan, A. (2012b). Value, authority and the open society. Some implications for digital and online archaeology, in C. Bonacchi (ed.) *Archaeology and Digital Communication: Towards Strategies of Public Engagement*: 1-14. London: Archetype.
- Bevan, A. (in press) The data deluge, *Antiquity*.
- Bevan, A., Li, X.J., Martínón-Torres, M., Green, S., Xia, Y., Zhao, K., Zhao, Z., Ma, S., Cao, W. and T. Rehren (2014). Computer vision, archaeological classification and China's terracotta warriors, *Journal of Archaeological Science*. 49: 249-254. <http://dx.doi.org/10.1016/j.jas.2014.05.014>.
- Bland, R. (2005). Rescuing our neglected heritage: the evolution of the Government's policy on Portable Antiquities and Treasure, *Cultural Trends* 14.4: 257-96.
- Bonacchi, C. (ed. 2012). *Archaeology and Digital Communication: Towards Strategies of Public Engagement*, London: Archetype.
- Bonacchi, C. (2014). Understanding the public experience of archaeology in the UK and Italy: a call for a sociological movement in Public Archaeology, *European Journal of Post-Classical Archaeologies* 4: 377-400.
- Brohan, P., Allan, R., Freeman, J.E., Waple, A.M., Wheeler, D., Wilkinson, C. and S. Woodruff (2009). Marine observations of old weather, *Bulletin of the American Meteorological Society* 90.2: 219-230.
- Causser, T., Tonra, J., and V. Wallace (2012). Transcription maximized; expense minimized? Crowdsourcing and editing The Collected Works of Jeremy Bentham, *Literary and Linguistic Computing* 27.2: 119-137.
- Ducke, B., Score, D., and J. Reeves (2011). Multiview 3D reconstruction of the archaeological site at Weymouth from image series, *Computers and Graphics* 35: 375-382.
- Dunn, S. and M. Hedges (2012). *Crowd-sourcing Scoping Study. Engaging the Crowd with Humanities Research*, Report for the UK Arts and Humanities Research Council Connected Communities Scheme. <http://crowds.cerch.kcl.ac.uk/wp-content/uploads/2012/12/Crowdsourcing-connected-communities.pdf>
- Haythornthwaite, C. (2009). Crowds and communities: light and heavyweight models of peer production, in C. Haythornthwaite and A. Gruzd (eds.) *Proceedings of the 42nd Hawaii International Conference on System Sciences*. Los Alamitos, CA: IEEE Computer Society. <https://www.ideals.uiuc.edu/handle/2142/9457>
- Hück, S. (2011). Be prepared for the coming "War for Co-Creators", *Open Business Council* blogpost (August 22, 2011). <http://www.openbusinesscouncil.org/2011/08/be-prepared-for-the-coming-war-for-co-creators/>
- James, M.R. and S. Robson (2012). Straightforward reconstruction of 3D surfaces and topography with a camera: Accuracy and geoscience application, *Journal of Geophysical Research* 117: F03017.
- Johnson, M. (2011). On the nature of empiricism in archaeology, *Journal of the Royal Anthropological Institute* 17: 764-87. URL: <http://dx.doi.org/10.1111/j.1467-9655.2011.01718.x>
- Kansa, E., Kansa, S. and E. Watrall (eds. 2011). *Archaeology 2.0: New Approaches to Communication and Collaboration*, Los Angeles: Cotsen Institute of Archaeology. URL: <http://www.escholarship.org/uc/item/1r6137tb>
- Lake, M. (2012). Open archaeology, *World Archaeology* 44.4: 471-8. <http://dx.doi.org/10.1080/00438243.2012.748521>

- Latour, B. 1990. Visualisation and cognition. Drawing things together, in M. Lynch and S. Woolgar (eds.) *Representation in Scientific Practice*: 19–68. Cambridge, MA: MIT Press.
- MacLeod, N. (2010). Alternative 2D and 3D form characterization approaches to the automated identification of biological species, in P.L. Nimis and R. Vignes Lebbe (eds.) *Tools for Identifying Biodiversity: Progress and Problems*: 225-229. Trieste: University of Trieste.
- Needham, S., Bronk Ramsey, C., Coombs, D., Cartwright, C. and P.B. Pettitt (1998). An independent chronology for British Bronze Age metalwork: the results of the Oxford Radiocarbon Accelerator programme, *The Archaeology Journal* 154: 55-107.
- Roy, H. E., Pocock, M. J. O., Preston, C. D., Roy, D. B., Savage, J., Tweddle, J.C. and Robinson, L.D. (2012). *Understanding Citizen Science & Environmental Monitoring*. Final Report on behalf of UK-EOF. NERC Centre for Ecology & Hydrology and Natural History Museum.
http://www.ceh.ac.uk/news/news_archive/documents/understandingcitizenscienceenvironmentalmonitoring_report_final.pdf
- Simon, N. (2010). *The Participatory Museum* (Museum 2.0). URL: <http://www.participatorymuseum.org/>
- Snavely, N., Seitz, S.M. and R. Szeliski (2008). Modeling the world from Internet photo collections, *International Journal of Computer Vision* 80: 189-210.
- Terras, M. (2010). Digital curiosities: resource creation via amateur digitisation, *Literary and Linguistic Computing* 25.4: 425-438.
- Verhoeven, G., Doneus, M., Briesec, C. and F. Vermeulen (2012). Mapping by matching: a computer vision-based approach to fast and accurate georeferencing of archaeological aerial photographs, *Journal of Archaeological Science* 39: 2060-2070.
- Wheat, R.E., Wang, Y., Byrnes, J.E. and J. Ranganathan (2013). Raising money for scientific research through crowdfunding, *Trends in Ecology and Evolution* 28.2: 71-72.