Archaeology Research Grant 2022

Jessica Smyth

Submission Date	Nov 26, 2022 7:54 AM
1. Title:	Dr
First name:	Jessica
Surname:	Smyth
2.	
3. Grant programme	Archaeology Research Grant
4. Year awarded	2022
5. Title of project	DigiHerd: digitally preserving Ireland's earliest and largest cattle assemblage

6. Summary of report (Minimum allowed 100 words)

From June to November 2022, the DigiHerd project undertook pilot X-ray CT scanning of a heavily fragmented Neolithic cattle skull (c. 200 fragments) excavated from an enclosure ditch at Kilshane, Co. Dublin. Representing nearly sixty individuals, the cattle remains from Kilshane is the largest Neolithic faunal assemblage in Ireland and one of the largest in Europe. It is an internationally significant reference collection, but its scientific potential and long-term conservation is threatened by the very fragmentary condition of the bone. A project team of archaeologists, soil scientists and computer scientists from the Schools of Archaeology, Agriculture & Food Science, and Computer Science at UCD aimed to re-assemble this complex 3-dimensional 'jigsaw puzzle' using the latest scanning technology alongside Machine Learning (ML) and other computational approaches. Initial findings from this pilot are promising, with a workflow for future scaled-up analysis worked out. The final machine learning reconstruction step is ongoing, with preliminary results expected in early 2023.

7. Please provide two appropriate images which can include photograph

8. Please outline the objectives of the project

DigiHerd aims to explore the potential for digital preservation of a Neolithic cattle assemblage via cutting-edge X-ray microCT and 3D surface scanning. Representing nearly sixty individuals, the cattle remains excavated from an enclosure site at Kilshane, Co. Dublin, in 2004 is the largest Neolithic assemblage in Ireland and one of the largest in Europe. It is an internationally significant reference collection, but its scientific potential and long-term conservation is threatened by the very fragmentary condition of the bone. A project team of archaeologists, soil scientists and computer scientists will aim to re-assemble this complex 3-dimensional 'jigsaw puzzle' using advanced scanning technology alongside Machine Learning (ML) and other computational approaches. With an ultimate goal of producing a digital archive of specimens that can be manipulated and studied virtually, this novel approach has huge potential for the future analysis of excavated skeletal remains, especially fragmented remains, as well as providing an important educational and scientific communication tool.

The Kilshane cattle bone is currently stored at the UCD School of Archaeology, where it is undergoing osteometric and isotope analysis as part of the IRC Laureate 'Passage Tomb People' project (www.passagetombpeople.com; PI: Jessica Smyth). The DigiHerd project is designed as a pilot project, complementing and running in parallel with the above research. We aim to employ non-destructive X-ray microCT scanning and surface scanning on a sub-set of the Kilshane material to ascertain feasibility and suitability of full-assemblage digital recording and archiving.

Specifically, we are targeting one the cattle skulls for scanning and modelling. The skulls are the most fragmented elements of this faunal assemblage and the most challenging to reconstruct digitally. They are thus thought to provide the best test of the scanning and computer modelling technology. Skulls are also the most recognisable part of the cattle skeleton and re-assembling digitally the fragmented skull will reinstate the visual impact lost during excavation.

Project team:

- Dr Jessica Smyth, School of Archaeology, UCD
- Dr Saoirse Tracy, School of Agriculture and Food Science, UCD; Director of UCD X-ray CT Facility.
- Dr. Stephen Kehoe – X-ray CT facility technical officer
- Dr Anthony Ventresque, School of Computer Science, UCD; Director of UCD Complex Software Laboratory.

9. Please describe the methodology used in conducting the research

Our target skull, sample #1059 (from deposit 5418 in the enclosure ditch 5001), comprised approximately 200 fragments (see Image 1).

DigiHerd originally aimed to use two image capture techniques side-by-side: microCT (X-ray) scanning and surface scanning. MicroCT scanning would take place at the UCD X-ray CT Facility, School of Agriculture and Food Science, with surface scanning taking place within the UCD School of Archaeology, using a handheld surface scanner (Einscan Hybrid LED and Infrared Light Source handheld colour 3D scanner). The digital outputs – the two scanning datasets - would then be processed in the UCD Complex Software Laboratory, in the School of Computer Science using Machine Learning (ML) to assemble the fragments using a computational 'best fit' approach. The rationale was that combining both 'internal' and 'external' image capture techniques would provide the maximum amount of bone reference points or 'landmarks' when digitally refitting the bone fragments. The variability of both internal bone structure and surface texture would be used to mathematically model which bone fragments fit to one another.

Work on DigiHerd commenced in early July 2023, following award acceptance and set-up at UCD. Project workflow was organised into five steps: sample pre-processing; selection of scanning technology; scanning and production of raw data; reconstruction and cleaning; machine learning. Pre-processing involved sorting the fragments from cattle skull S1059 into broad anatomical areas, aided by a colleague from UCD Zoology, so we could then sequentially scan clusters of spatially related bone (Image 1). This clustering step aimed to reduce the search space during the final machine learning/reconstruction step. The scanning technology selection step aimed to establish methods to obtain as high quality a scan as possible, in a way that would be consistent and quick to set up. This involved trials using a Microsoft Kinnect, an Einscan Hybrid LED and Infrared Light Source handheld colour 3D scanner, and the X-ray scanner at UCD Rosemount. The scanning/production of raw data step involved the X-ray scanner (see below), with each bone fragment wrapped in low-absorption materials (small plastic sample bag and then into a Styrofoam holder) before being scanned (Image 2). Once the images had been captured for each fragment, a data reconstruction and cleaning step assembled these X-ray image 'slices' into a 3-D form and removed image 'noise'. A final machine learning step involved investigating the best ways to re-assemble the scanned fragments or 3-D objects into a cattle skull. The file scans were uploaded to a high-performance server managed by UCD, allowing us to perform complex calculations extremely quickly and much faster than using personal machines.

Team member Dr Ventresque secured additional funding from the UCD Institute for Discovery for the data processing steps, which were carried out by postgraduate and postdoctoral members of his team.

10. Please outline the findings of your research and/or milestones achieved

Scanning technology – trials with different scanning technologies established that surface scanning was not compatible with our aims of high scan quality, consistency and quick set-up. As we needed a full 360 degree surface scan, the bone fragment would need to be suspended in some way and setting this up for each scan, particularly for very small fragments, would be quite time-consuming and counterproductive. Additionally, the surface scanner is typically used for larger objects and we observed some of our smaller bone fragments weren't getting picked up by the scanner. Finally, the fact that only the surface was scanned meant that we would be losing valuable information about the inside of the bones, which is crucial to figuring out which of these fit together best. We established that the X-ray scanning technology on its own could provide extremely high-quality and consistent imaging in a very short time period (each scan lasting just 9 minutes).

Data reconstruction and cleaning - The X-ray scanner rotates each bone fragment and captures roughly 1,000 images of it from every angle in .tiff format. The result of this is called a .tiff stack, similar to a flipbook that shows the cross-section of the bone fragment with each frame. To turn these images into a 3-D object we used VG Studio Max software, top-of-the-line software for tiff stack object reconstruction. We experimented with using open source software like ImageJ but the results were not as good. The resulting mesh was saved as an .stl file, a file format we found was the easiest to work with, as it can be opened in any 3D model editor such as Meshroom or Blender. We established that the quickest way to clean the mesh of any traces of the plastic sample bags inadvertently captured in the scans, was to load the mesh in Meshroom and select the largest connected piece, i.e. the bone, then select all polygons not connected to this bone, i.e. the irregularities, before deleting them.

Machine Learning – this comprised the largest and most unknown part of the DigiHerd pilot, and while we are still working towards the final skull reconstruction we have made good progress in the preparatory steps. Before performing any machine learning complex calculations, we had to determine the best techniques to use in reconstructing the skull. The first step was grouping the bone fragments further, to create a smaller number of objects on which to operate (clustering), and then using this smaller search space to find fragments that were most likely connected to another fragment (simulated annealing). To further group the objects we explored variables such as size, porosity, distribution (using binary encoding) and point cloud density. The preliminary findings of this clustering step are promising and we hope to have outputs and a publication on this in Q2 2023 (see below).

11. a) Please provide details of the dissemination of the outcomes from this project (inc. publications, presentations, outreach, media etc.) including details of any social media/web platforms used to publicise this project	There has been no external dissemination of the project outcomes in the period July to November 2023, as project results are only emerging. See below for planned dissemination into 2023. Postgraduate team members Vlad Rakhmanin and Elise Brard (School of Computer Science, UCD) gave a presentation on the DigiHerd project as part of the UCD Complex Software Lab seminar series in October 2022.
b) No. of Academic Papers/articles published:	0
c) No. of Lectures given/outreach events involved in:	1
d) Media Coverage (article in local newspaper, feature on University website etc.):	0
e) How will you continue to communicate the results of your project and what are your publication plans?	A journal article on the results of the DigiHerd project is currently being drafted. Target journals being considered include Antiquity, Journal of Advanced Manufacturing Technology, and Journal of Archaeological Science. Publicity activities are in train. Details of the project will be sent to the UCD press office which pushes press releases to various channels, e.g. Kevin O'Sullivan, science editor in the Irish Times. We have reached out to UCD science communicator Dr Shane Bergin, who has flagged our project with Jonathan McCrea who does the Newstalk science show 'Futureproof'. We plan to present project results at a number of outreach events, e.g. UCD Festival (June 2023), Heritage Week (August 2023), Science Week (November 2023). We will approach the Education department of the National Museum of Ireland and Natural History Museum about scheduling a lunchtime lecture in 2023.

15. How did the award enhance your professional development (e.g. in terms of specific opportunities, opportunities for enhancing skills, collaborations with others etc.)?

The DigiHerd pilot project has enabled us to assemble a very talented interdisciplinary team to address a very timely issue in Archaeology/cultural heritage. It will undoubtedly lead to more scaled-up projects in this vein (see below)

16. What plans (if any) do you have to further your proposal/project?

The final step of machine learning is the reconstruction itself and will continue with DigiHerd in the coming months. Our next objective is to use Operations Research techniques, designed for scenarios where computational complexity is high and the search can be driven towards a good solution. The DigiHerd pilot has enabled us to build a team with all the domain knowledge we need to make it successful. The first technique we are currently evaluating is simulated annealing - an algorithm that generates the most optimal configuration of something. In this case, we want to see the optimal way that each of these fragments can be grouped together into a single skull. We will perform this algorithm first on the bones within each individual clustered group and use the resulting combinations to check how these can fit in with the other groups. Once this is completed, we will perform the simulated annealing step and hopefully see some positive results. We are also considering how we can display our results, regardless of how the machine learning algorithm works, possibly a 'digital museum' or 'digital laboratory' that will allow users to view and rotate the bones in an online library.

The scanning and modelling completed as part of the DigiHerd pilot provides important baseline data for future work. Projected costs and timings of scanning and image analysis/processing will be used to prepare a scaled-up research proposal to a suitable funding scheme. Relevant schemes include the Strand 1L (INSTAR+) and Strand 2A of the IRC COALESCE (Collaborative Alliances for Societal Challenges) fund. Strand 1L is supported by the National Monuments Service of the Department of Housing, Local Government and Heritage in partnership with the Heritage Council, with awards up to €220,000. Strand 2A looks for interdisciplinary research addressing national or global societal challenges led by an AHSS PI with STEM co-PI, again with awards up to €220,000.

Another potential funding target is via Cluster 2 (Culture, Creativity and Inclusive society) of the EU's key funding programme Horizon Europe. The cluster includes the aim of safeguarding cultural heritage and the 'digitalisation of heritage' is set to be part of 2023 thematic areas/calls. With the imaging dataset from DigiHerd we will be well placed to develop a European-level consortium targeting this sort of vulnerable and under-utilised archaeological material and the development of an open source, user friendly and cost-efficient solution for 3D recording.