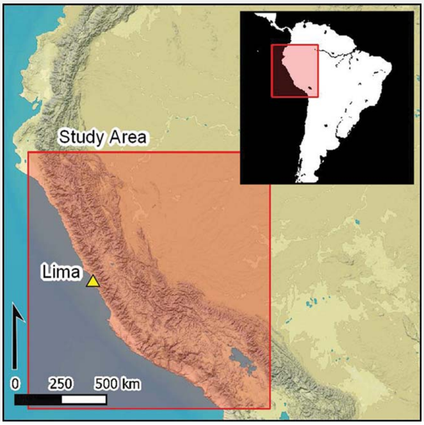
Antiquity 2024 Vol. 98 (397): 155–171
https://doi.org/10.15184/aqy.2023.177

Research Article: GeoPACHA

Large-scale, collaborative imagery survey in archaeology: the Geospatial Platform for Andean Culture, History and Archaeology (GeoPACHA)

Steven A. Wernke[1,*], Parker Van Valkenburgh[2,*], James Zimmer-Dauphinee[1], Bethany Whitlock[2], Giles Spence Morrow[1], Ryan Smith[3], Douglas Smit[4], Grecia Roque Ortega[5], Kevin Ricci Jara[5], Daniel Plekhov[6], Gabriela Oré Menéndez[7], Scotti Norman[8], Giancarlo Marcone[9], Manuel Mamani Calloapaza[10], Lauren Kohut[11], Carla Hernández Garavito[12], Sofia Chacaltana-Cortez[13] & Elizabeth Arkush[3]

[1] Department of Anthropology, Vanderbilt University, Nashville, USA
[2] Department of Anthropology, Brown University, Providence, USA
[3] Department of Anthropology, University of Pittsburgh, USA
[4] Department of Anthropology, University of North Carolina at Chapel Hill, USA
[5] Facultad de Ciencias Sociales – Arqueología, Universidad Nacional Mayor de San Marcos, Lima, Peru
[6] Anthropology Department, Portland State University, Portland, USA
[7] Department of Anthropology, University of Nevada, Las Vegas, USA
[8] Department of Sociology and Anthropology, Warren Wilson College, Asheville, USA
[9] Dirección de Humanidades Artes y Ciencias Sociales,Universidad de Ingenieria y Tecnología, Lima, Peru
[10] Facultad de Ciencias Histórico Sociales, Universidad Nacional de San Agustín, Arequipa, Peru
[11] Department of Environmental Studies, Winthrop University, Rock Hill, USA
[12] Anthropology Department, University of California, Santa Cruz, USA
[13] Programa de Humanidades, Universidad Antonio Ruiz de Montoya, Lima, Peru
* Authors for correspondence ✉ s.wernke@vanderbilt.edu & parker_vanvalkenburgh@brown.edu



Imagery-based survey is capable of producing archaeological datasets that complement those collected through field-based survey methods, widening the scope of analysis beyond regions. The Geospatial Platform for Andean Culture, History and Archaeology (GeoPACHA) enables systematic registry of imagery survey data through a 'federated' approach. Using GeoPACHA, teams pursue problem-specific research questions through a common data schema and interface that allows for inter-project comparisons, analyses and syntheses. The authors present an overview of the platform's rationale and functionality, as well as a summary of results from the first survey

Received: 3 October 2022; Revised: 18 April 2023; Accepted: 6 June 2023

campaign, which was carried out by six projects distributed across the central Andes, five of which are represented here.

Keywords: South America, Andes, satellite survey, settlement pattern analysis, big data, large-scale imagery survey, QGIS

Scalar challenges in archaeology

One of archaeology's most substantial challenges is aligning the scales of our datasets with those of the social worlds that we seek to study. At the smaller end of the scalar spectrum, archaeologists harness an ever-expanding range of scientific techniques to conduct detailed analyses of artefacts and sites, enriching understanding of human social experience and pushing back against the generalities of grand historical narratives (e.g. Mills & Walker 2008; Hegmon 2016; Roddick & Stahl 2016; Supernant *et al.* 2020). The discipline has also had great success working at the scale of localities and regions, through pedestrian survey projects and settlement pattern studies (e.g. Johnson 1977; Banning 2002; Cherry 2003; Kantner 2008; Drennan *et al.* 2015; Alcock & Cherry 2016). But conventional archaeological methods and protocols are often ill-suited for collecting systematic data at interregional and continental scales.

The largest pedestrian surveys—requiring many years of effort by large research teams—cover, at most, a few thousand square kilometres and tend to employ idiosyncratic classificatory systems that hinder inter-survey comparisons (Daniels 1970; Sanders 1970; Sanders *et al.* 1979; Adams 1981; Blanton *et al.* 1981, 1999; Barker *et al.* 1996; Bauer & Covey 2002; Bewley *et al.* 2016). Moreover, where survey data registries can be reconciled and aggregated, their combined distributions do not generally constitute systematic samples of large study areas. Instead, they represent targeted samples whose locations are influenced by such factors as contemporary land cover, national funding priorities, convenience, regulatory frameworks and the research interests of individual survey projects. Consequently, characterisations of interregional trends often end up resembling scaled-up versions of localised observations, and we have struggled to produce analyses of broader phenomena—such as continental-scale demographics, large-scale societal responses to environmental change and the political economies of expansive polities—with the same rigour that we would expect from archaeological studies conducted at the scale of sites, localities and regions.

In the temporal dimension, we face a corollary issue. While archaeology is uniquely equipped to produce knowledge of the deep past and to chart change in the long-term (Perreault 2019), the diversity of both recording standards (as they vary across projects and regions) and of the archaeological record itself (as it tends to become sparser and less accessible with age) often impede the aggregation of sufficient data to chart diachronic trends in rigorous fashion (Kintigh & Altschul 2010; Spielmann & Kintigh 2011; Altschul *et al.* 2018). Thus, just as scattered survey and excavation results must be pulled together to discuss continental-scale variation, archaeologists must also contend with patchy temporal coverage to map out change over time. These difficulties are compounded by the increasing abundance and richness of archaeological data.

https://doi.org/10.15184/aqy.2023.177 Published online by Cambridge University Press

Archaeology's scalar challenges are formidable, but systematic, large-scale research is vital for the future of the field, for at least two reasons. It is not so much that—as Perreault (2019) contends—archaeological data are inherently better suited for addressing long-term or large-scale research questions; rather, such 'big' archaeology is crucial, first, because it contributes to a diverse array of mutually enriching approaches. Just as highly localised research is essential for recording lived experiences that are often missing from expansive studies, large-scale, comparative research provides critical information for making sense of variation observed in smaller-scale inquiries. Archaeologists already appreciate this complementarity, but we lack access to systematic, continuous data collected at large scales. Second, working beyond the 'local' and the short term is also vital because the social and political horizons of populations are more expansive than small spatial and temporal scales. Like modern subjects, people in the past understood and acted in their worlds through multiscalar and long-term perspectives and were enmeshed in multiscalar social, natural and temporal processes.

Our desire to address these issues in the Andean region are what led to the development of GeoPACHA. GeoPACHA is a geospatial webapp built with an open-source software stack that is designed to enable diverse research teams to pursue large-scale, project-specific archaeological research questions. It serves high-resolution satellite and historical aerial imagery, allows users to tag features of interest, and provides editorial tools that enable careful tracking of survey coverage and data quality. Attribute data are recorded in a central PostgreSQL/POSTGIS database. Like some other imagery survey projects, GeoPACHA is designed to enable collaboration among team members spread across the globe. Unlike crowd-sourcing platforms, however, it is intended to facilitate survey by trained researchers, supervised by domain experts conducting problem-oriented research. While users work within a shared framework, each is the member of a research team pursuing project-specific research questions.

In the first deployment of GeoPACHA (2020–21), users tagged areas of archaeological interest ('loci') based on research questions established by project supervisors ('regional editors'), who directed research in each survey zone. Locus identifications and attributes were then reviewed twice—first by the regional editors, then by 'general editors' (Wernke and VanValkenburgh). Large-scale imagery survey through GeoPACHA enabled six teams to pursue distinct research questions in different areas of the Andes—the northern montaña and highlands, north coast, central coast, central highlands and southern highlands of Peru, and the circum-Titicaca Basin of Peru and Bolivia (Figure 1). Six of these studies (including this article) are presented in *Antiquity* (Arkush *et al.* 2023; Marcone *et al.* 2023; Spence Morrow *et al.* 2023; Whitlock *et al.* 2023; Zimmer-Dauphinee *et al.* 2023).

This article provides an overview of these results and the potential of large-scale archaeological imagery survey in the central Andes and beyond. We describe the functionality of GeoPACHA and discuss the prospects and challenges of its federated, peer-reviewed framework. We contend that, while the platform (like all large-scale imagery survey projects) is not well-suited for addressing certain research questions and is not useful in all landscape types, the continuous coverage enabled by GeoPACHA has already significantly enhanced our understanding of archaeological settlement patterns and landscapes in the central Andes. Equally importantly, project results are already generating new questions that might be addressed through future field research.

https://doi.org/10.15184/aqy.2023.177 Published online by Cambridge University Press

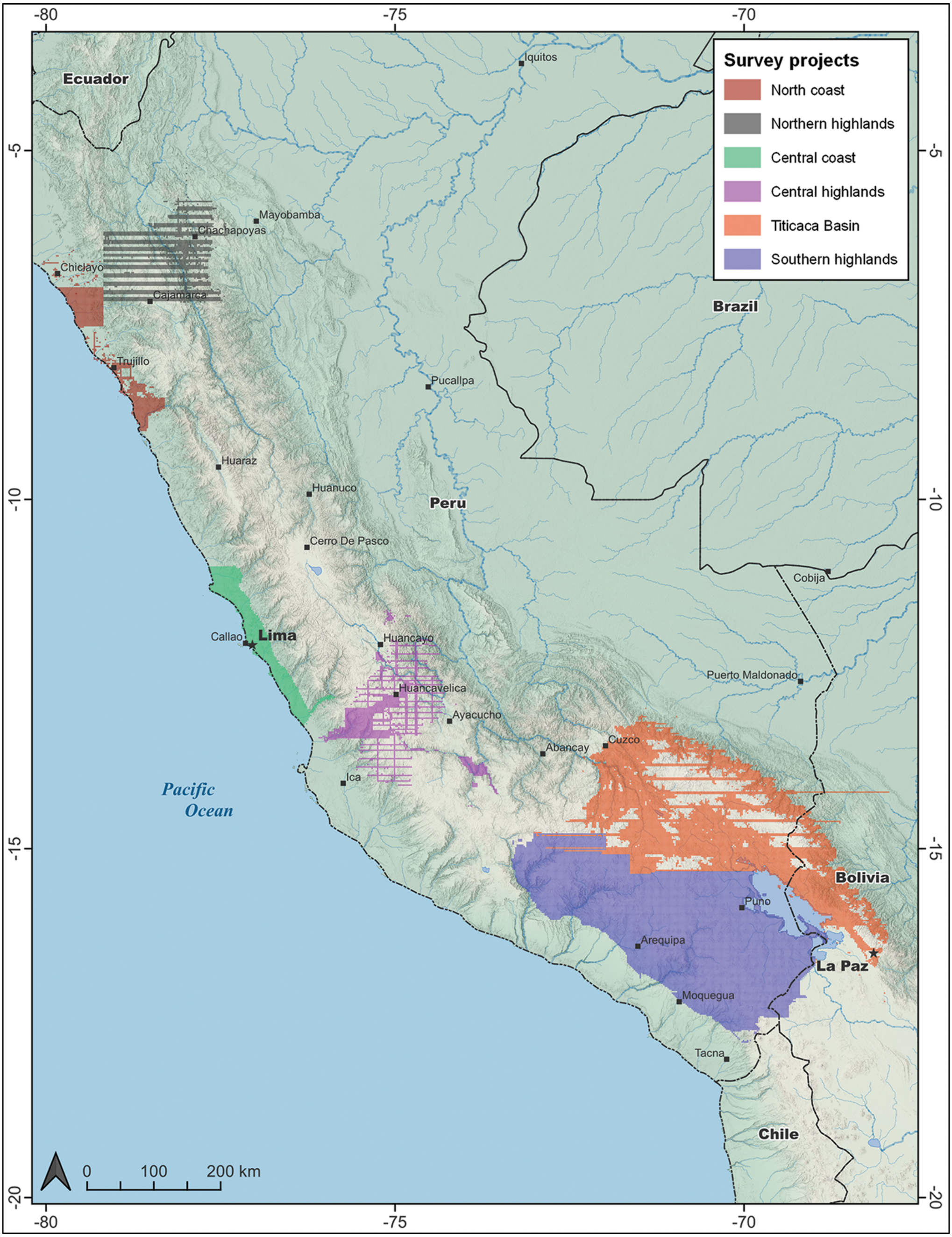


Figure 1. GeoPACHA survey project areas (figure by S.A. Wernke).

https://doi.org/10.15184/aqy.2023.177 Published online by Cambridge University Press

Problems of scale: linked open data repositories and imagery surveys

To date, efforts to overcome archaeology's problems of scale have concentrated on two approaches: linked open data repositories and large-scale imagery survey. The former include the Digital Archaeology Record (Spielmann & Kintigh 2011; Alleen-Willems 2012; McManamon *et al.* 2017), Open Context (Kansa & Kansa 2007; Kansa *et al.* 2007; Kansa 2012), the Digital Index of North American Archaeology (Wells *et al.* 2014; Kansa *et al.* 2018), and the Archaeology Data Service. In Peru (the core GeoPACHA coverage area), the Sistema de Información Geográfica de Arqueología by the Ministry of Culture of Peru acts as a growing (but not yet linked or open-source) clearinghouse for some archaeological project data. These efforts have greatly improved access to field data that were previously stored in disparate silos, and they have made it possible to conduct analyses at larger scales by resolving differences among bespoke data schema (e.g. Atici *et al.* 2013; Anderson *et al.* 2017). But as the archived datasets are produced by individual archaeological projects, linked open repositories cannot themselves overcome the sampling biases of previous field research coverage.

A principal and complementary contribution of large-scale imagery survey is that it facilitates the collection of new archaeological datasets that do not inherit these legacies. Archaeologists have been quick to leverage high-resolution satellite imagery to map archaeological features, especially those in areas with sparse land cover (Ur 2006; Parcak 2009). Data collection protocols tend to follow three models: 1) citizen science; 2) what Casana (2014: 226) calls "brute force" survey; and 3) automated detection. Citizen science projects, which train non-specialists to tag archaeological features *en masse*, include Parcak's (2019) GlobalXplorer project and Lin and colleagues' (2014) search for Ghengis Khan's tomb. Brute force surveys, in which smaller teams with domain-specific expertise visually scan satellite imagery and tag features, include Casana's own CORONA Atlas (Casana & Cothren 2013), the Endangered Archaeology in the Middle East and North Africa project (Bewley *et al.* 2016) and Caucasus Heritage Watch (Caucasus Heritage Watch 2022). Finally, automated approaches include both probabilistic modelling of sites and soils (e.g. Menze & Ur 2012) and more recent deep learning approaches that appear to significantly improve feature detection (e.g. Soroush *et al.* 2020; Bickler 2021; Cao *et al.* 2021). In this special section, Zimmer-Dauphinee and colleagues (2023) report on a human-machine teaming approach that shows promise for further upscaling of GeoPACHA through semi-automated locus detection.

Each of these approaches has both benefits and limitations. Crowdsourcing broadens participation and facilitates collection of massive datasets, but it can suffer from data quality issues and the translation of broad goals into specific research contributions. For example, GlobalXplorer's survey of Peru covered about 20 per cent of the country (150 000km^2) and registered 19 084 sites with the help of over 70 000 remote volunteers (GlobalXplorer 2018), but it has yet to lead to scientific publications. Lin and colleagues' crowdsourced efforts to locate the tomb of Genghis Khan drew upon over 10 000 volunteers, some 30 000 hours of effort, and generated 2.3 million feature categorisations (Lin *et al.* 2014). These efforts enabled identification of 55 ground-truthed archaeological sites, but no candidate for the tomb itself (Lin *et al.* 2014; Casana 2020).

https://doi.org/10.15184/aqy.2023.177 Published online by Cambridge University Press

Brute-force survey has produced high quality data that have broadened archaeological perspectives to interregional scales, especially in the Near East (Casana 2014; Casana & Panahipour 2014; Casana 2015). The CORONA Atlas has surveyed 300 000km^2 from eastern Egypt through Mesopotamia and documented over 14 000 sites (Casana & Cothren 2013; Casana 2014). Of these, about 10 000 were previously undocumented—both because the imagery survey encompassed vast areas that had never been systematically surveyed and because the historical CORONA satellite imagery used in the project enabled detection of sites since destroyed (Casana 2014; Casana & Panahipour 2014; Casana 2015). These are spectacular results, and they prove that large-scale research need not be carried out by massive teams nor using automated methods. As the term implies, however, brute force survey requires research teams to dedicate large outlays of time and often monotonous effort to cover areas mostly devoid of visible archaeological remains.

The promise of archaeological imagery survey thus lies in its potential to expand geographic frames of reference, generate continuous datasets and (when based on historical imagery) to document features and sites that today have been destroyed, degraded or obscured. At the same time, it poses epistemological, methodological and ethical questions that need to be addressed. Working at interregional scales requires simplified and generalised data schema that may not capture all dimensions of variation. Additionally, because not all sites are visible in aerial and satellite imagery, there is a non-trivial false negative problem in all forms of imagery-based survey. (It is worth noting, however, that this problem is common to archaeological data collection, due to selective preservation and visibility). Finally, the chronology of features identified in satellite imagery can only be estimated where these features have temporally diagnostic forms; identified distributions of other feature types represent cumulative records (i.e. palimpsests) rather than occupations dating to discrete periods.

Fortunately, many of these biases can be modelled. Landscapes can be subdivided based on surface visibility and geomorphology, to estimate where features are likely to be undersampled. Likewise, we can simulate how sites of certain types and ages (for example, older sites) might be underrepresented in imagery survey datasets (Contreras & Meadows 2014). Yet, like field research, imagery survey inevitably entails compromises between coverage and intensity. If excavation affords relatively thick descriptions of archaeological sites, and field survey produces thinner data over larger areas, then imagery survey generates perhaps the thinnest data of all. To draw an analogy from the digital humanities, imagery survey is akin to distant reading (Moretti 2013); its hermeneutics are complementary to those of field-based archaeology, as distant reading is complementary to close reading. Each method probes different dimensions of complex, underlying phenomena. We thus see the value of imagery survey as providing a valuable new layer or overlay of continuous archaeological distributional data at scales unobtainable through field-based methods.

For these reasons, we resist framing imagery survey as anything other than just another tool in the archaeologist's toolbox. It is no substitute for fieldwork and provides no reasonable means by which we might map all archaeological sites across the globe. It simply provides us with new (productive, but also partial and highly situated) vantages. Because popular media often resort to techno-utopian tropes to describe digital archaeological projects, it is incumbent that we continually ground our work by explicating its specific affordances and

https://doi.org/10.15184/aqy.2023.177 Published online by Cambridge University Press

limitations, while mitigating against the possibility that publishing large-scale datasets will facilitate site destruction and/or unauthorised surveillance. While there are no easy solutions to these problems, epistemic humility and collaborations with host communities and national heritage institutions are essential starting points.

GeoPACHA: platform design and survey results

We designed GeoPACHA's collaborative framework to address the above-mentioned challenges and prospects. GeoPACHA is a 'federated' platform: it uses a common data ontology and schema to enable observational and analytical commensurability across survey projects, while also being extensible and customisable to accommodate diverse research questions and designs (in this sense, it draws inspiration from the FAIMS project; Ross *et al.* 2013). The federated concept was intended to facilitate problem-based data collection, to be carried out by archaeologists with field experience in their respective imagery survey zones, while also employing common attributes and vocabularies so that datasets could be merged across projects where so desired.

Development of the webapp began with the codebase of another well-known imagery survey platform—the CORONA Atlas, developed by Jesse Casana and colleagues (Casana & Cothren 2013). GeoPACHA was initially built on an open-source software stack, with MySQL handling the database backend and PHP scripting controlling the user interface, experience and permissions within the CodeIgniter framework. Following a workshop at Vanderbilt University in 2019, in which project members outlined their goals for imagery survey, we adapted the existing codebase to our system needs. The first survey campaign was conducted using a version of the webapp built with this revised codebase.

Following the first survey campaign, we converted the MySQL database into a PostgreSQL/POSTGIS database so that each survey team could make further edits and amendments while conducting analysis via QGIS, the most widely used open-source desktop GIS application. This latest version preserves the structure, version control functions and user privileges of the original database. The backend of the webapp was also converted to point to the PostgreSQL/POSTGIS database, so that users can now connect to a single canonical database either via the webapp or QGIS. Given the sensitivity of some site locational data, access to GeoPACHA is restricted to registered users. We are now designing a repository of survey results to be accessible via registered users through Open Context.

The GeoPACHA webapp enables the user to toggle between several imagery sources (including Google, Bing, ESRI and Mapbox, as well as a 0.25m-resolution orthomosaic of the Colca Valley derived from photographs from the 1931 Shippee-Johnson Peruvian Expedition), place points where an archaeological locus is detected, and fill out an attribute form. The attribute data schema is a key element of the federated concept of GeoPACHA, allowing different projects to add specialized fields to address certain research questions while maintaining a common core that facilitates aggregation and analysis across all survey projects. Survey coverage is tracked using a tiered grid system (described below). Once a locus is recorded by a surveyor, the data are passed to a regional editor for review in a separate interface in which surveyors' initial locus identifications are listed. Regional editors then review each locus identification to accept or reject them, while also reviewing and editing their attribute data, as

https://doi.org/10.15184/aqy.2023.177 Published online by Cambridge University Press

necessary. Once a locus is approved by a regional editor, it is passed on to the general editors for review through the same interface. General editors then make final reviews of locus identifications and attributes, and approved loci are committed to the canonical database. GeoPACHA thus integrates two levels of peer-review into its design.

Survey coverage tracking is achieved via a grid-based tessellation over the survey areas. As surveyors zoom in to imagery within the webapp, grids appear at three scales: 0.02° (about 2 × 2km), 0.01° (about 1 × 1km), and 0.005° (about 0.5 × 0.5km). Thus, a given 2 × 2km cell is composed of four 1 × 1km and sixteen 0.5 × 0.5km cells. Surveyors, who are trained and co-ordinated by regional editors, then zoom in to imagery until a minimal (0.5 × 0.5km) grid cell fills their screen; they then visually scan it by systematically moving their eyes up and down in transects and are encouraged to zoom in to further investigate possible loci. Where loci are identified, surveyors record attribute information, including locus type, number of structures, extent and level visibility, as well as confidence indices. After all features in a given 0.5 × 0.5km cell have been investigated and tagged with appropriate attribute data, the surveyor moves on to the next one. When all sixteen 0.5 × 0.5km cells within a 2 × 2km cell are completed, the surveyor marks the encompassing 2 × 2km cell as complete. Regional editors can review these tagged cells before approving and sending them on to the general editors, or sending the cell back to the survey team for continued review.

To accommodate regional editors' diverse research objectives, we chose an intentionally capacious concept as the atomic unit of data registry: the locus. In our usage, a 'locus' refers to any discrete archaeological feature or set of features, with a threshold distance of 100m from the nearest other identifiable feature or set of features. That is to say, the project data schema is agnostic with regard to defining specific sites or settlements (Dunnell 1992; for recent discussion of this issue in relation to big digital archaeology, see McCoy 2020). The platform thus affords registry of landscape complexes, features or settlements as defined by participating projects. A locus could be a relict terrace complex, a settlement, a fortification or any other set of archaeological remains visible in imagery. Attributes are organised into nested fields with controlled vocabularies (via foreign key constraints in the PostgreSQL database). Thus, for instance, a complex of stone-faced terraces would be identified as a locus of type 'agro-pastoral infrastructure', with subtype 'stone faced terracing'. However, because not all regional editors were addressing research questions that were related to terraces, not all projects recorded their locations. Projects could opt to record locus areas using an area measurement tool in the platform interface, but locus boundary polygons were not stored as part of the project database because we reasoned that it would be of limited utility, while significantly hindering survey coverage.

Following the federated concept, research agendas for GeoPACHA projects were defined and pursued independently, but designed in consultation with the general editors to ensure that the platform could accommodate their needs. While some surveys registered all visible loci, others targeted a narrower range of locus types. For instance, the Titicaca Basin survey focused on hilltop fortifications (*pukaras*) dating to the Late Intermediate Period (AD 1000–1450) and Late Horizon (AD 1450–1532). In contrast, the adjacent southern highlands survey sought to record all visible remains. Yet because the two survey projects used the same data schema through GeoPACHA, the pukara identifications from the southern highlands zone could be combined with those of the Titicaca Basin survey, thereby greatly expanding the scope of systematic pukara registry (see Arkush *et al.* 2023).

https://doi.org/10.15184/aqy.2023.177 Published online by Cambridge University Press

The six initial survey projects covered a combined total of 179 427km^2 and registered a total of 38 753 archaeological loci (Figure 2, Table 1). The survey campaign ran from 15 January 2020 to 10 July 2021 and was then followed by spot checks, editing and data review.

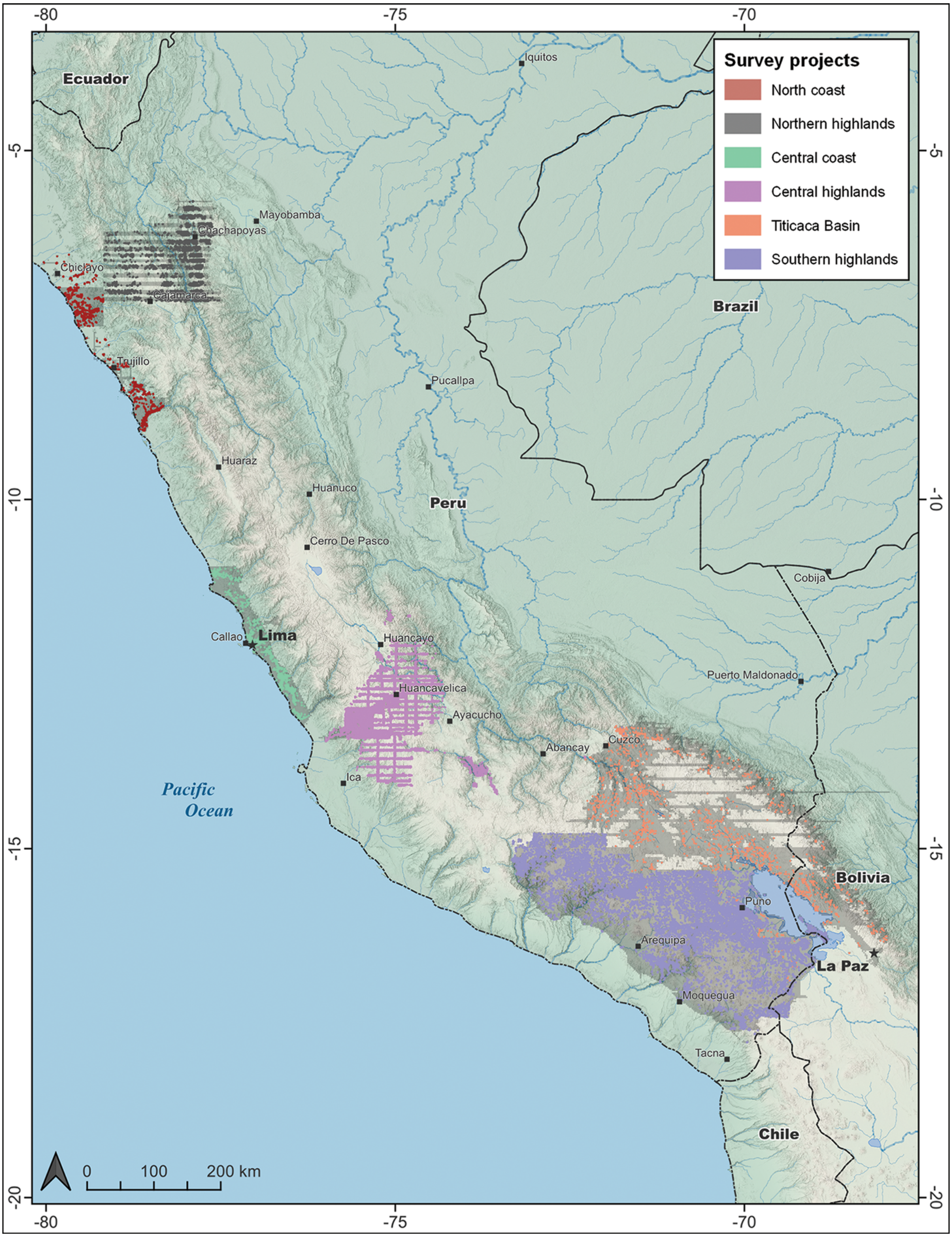


Figure 2. GeoPACHA loci registered, by survey project (figure by S.A. Wernke).

https://doi.org/10.15184/aqy.2023.177 Published online by Cambridge University Press

Table 1. Area covered by each survey project in the first survey campaign.

Survey project	Area (km^2)
Central coast	8451
Southern highlands/Titicaca Basin	58 857
Central highlands	12 199
Southwestern highlands	78 372
North coast	7524
Northern highlands	14 023
Total	**179 427**

The campaign's coincidence with the onset of the SARS-CoV-2 pandemic was of course unexpected, yet the pandemic did come to shape our work. We had initially planned for the survey to last only 12 months, but as the first full year of the pandemic set in and it became clear that conducting fieldwork would continue to be impractical, we extended the project period. For two doctoral students, it provided a vital means of collecting dissertation research data (Whitlock *et al.* 2023; Zimmer-Dauphinee *et al.* 2023); for others, it provided a means of conducting remote work. The platform made it possible to build year-round research projects that were international and inclusive, by enabling project members to work together on a virtual platform that did not require the ability to traverse difficult terrain on foot. In this first survey campaign, GeoPACHA teams were composed of 54 members from several countries, from professors and professionals to undergraduate students, with regional experts from Peru, Canada and the United States. Table 2 presents a summary of loci by type and survey project.

Discussion and conclusion

The articles that follow in this special section present analyses of data from our first survey campaign, as well as discussions of survey project rationales and designs. Each of these projects pursued distinct research agendas tailored to the affordances and limitations of large-scale imagery survey. Given their diversity, we will not attempt synthesis here, but one general insight that emerges is the highly uneven distribution of loci across Andean landscapes.

For example, the survey project in the southern Peruvian highlands (see Arkush *et al.* 2023) recorded 14 718 loci in a 78 372km^2 area; joining these loci to the finest grid used for guiding survey coverage (composed of 0.5 × 0.5km cells) shows that only 4.8 per cent of the grid cells have visible archaeological traces (Figure 3). Even adding in areas of terracing and other field systems that continue to be cultivated in the present (many of which are likely to have been cultivated in the past), archaeological loci are still visible in only 16 per cent of grid cells.

This pattern appears to be meaningfully related to the distribution of landforms and resources within the southern highlands survey area. Despite the general perception that human populations were ubiquitous in the Andes and that every valley contains terracing (e.g. Stanish 1987: 337), there are vast expanses of the highlands where no signs of

https://doi.org/10.15184/aqy.2023.177 Published online by Cambridge University Press

Table 2. Locus types registered within each survey zone

Locus type	North highlands	North coast	Central highlands	Central coast	Titicaca Basin	Southern highlands	Total
Settlement	211	452	3111	244	390	7878	**12 286**
Agropastoral infrastructure	4950	220	10 548	388	50	6230	**22 386**
Ambiguous	34	500	647	228	54	259	**1722**
Pukara	16	131	31	6	1030	211	**1425**
Path	1	32	5	8	30	99	**175**
Unclassified	13	287	27	157	5	23	**512**
Ritual/ ceremonial	–	101	–	29	21	11	**162**
Mortuary	–	24	–	17	9	7	**57**
Geoglyph	–	25	–	3	–	–	**28**
Total	**5225**	**1772**	**14 369**	**1080**	**1589**	**14 718**	**38 753**

human habitation or landscape modification are visible in contemporary satellite imagery. Because many of these areas are also not currently inhabited and are difficult to reach, they are also places where pedestrian surveys are less likely to be conducted. As a result, these areas tend to be excluded from the survey datasets we use to understand ancient settlement patterns and demographics. The result is that our current models of settlement distribution are biased in favour of densely inhabited areas—perhaps so much so that we have not been able to fully appreciate the range of factors that have led Andean peoples to live where they do. In their contributions to the GeoPACHA articles, Marcone *et al.* (2023) and Spence Morrow *et al.* (2023) explore how modern settlement patterns and environmental conditions have impacted archaeological data collection, and they use GeoPACHA to provide alternative vantage points.

To extend these implications further, one aim shared among GeoPACHA research projects has been understanding relationships between pastoralist and agriculturalist settlements, through the identification of ancient corrals and agricultural fields. While a thorough analysis of the resulting data is beyond the scope of this article, there are strong indicators that the distribution of these locus types in the southern highlands is not driven solely by the distribution of resources. Rather there seem to be strong and durable social links driving the distribution of pastoralist populations in relation to agriculturalist populations, with particularly tight coupling between valley sites found at 3200–3800m above sea level and pastoral sites found at 4000–4500m above sea level. These patterns are evident in many (but not all) portions of the survey area that fall within the given elevation bands. Without systematic large-scale imagery survey coverage, not only would we not have identified this pattern, but we might also have not considered the possibility that it could reflect something other than environmental factors. Though we can only gesture towards these patterns in this overview piece, they exemplify the kind of cumulative, long-term and inter-regional scale distributional view uniquely enabled by imagery survey that we advocate for as a complement to field-based research.

https://doi.org/10.15184/aqy.2023.177 Published online by Cambridge University Press

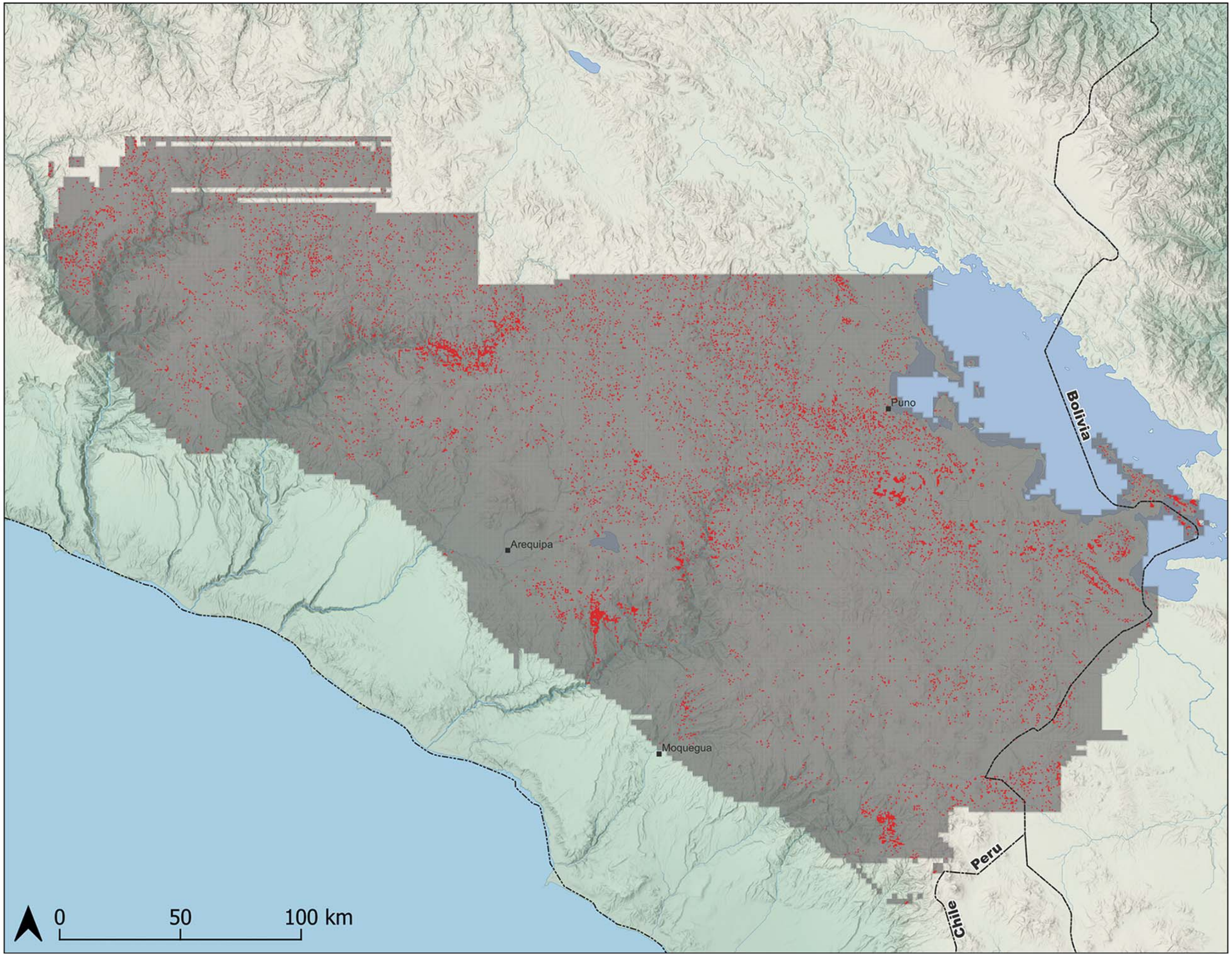


Figure 3. Minimal grid cells with loci present, southern highlands survey zone (figure by S.A. Wernke).

At the same time, the fact that such a high percentage of the smallest (0.5 × 0.5km) survey grid cells contained no loci posed real methodological challenges—not least of which was observation fatigue. Our surveys do not register full censuses of loci visible in the imagery used, though we are confident they represent a very large and representative proportion of them. In their contribution to this special section, Zimmer-Dauphinee and colleagues discuss these issues in their development of automated feature detection using machine learning models and compare them to the GeoPACHA human-tagged dataset. It is in large measure due to this issue of general occupational sparseness that we are advancing deep learning approaches. Our next stages of development thus seek to synergise artificial intelligence (AI) and human expertise by leveraging the large dataset of human-tagged archaeological features from this stage of the GeoPACHA imagery survey to further refine the deep learning models we have already developed, deploying those models for autonomous archaeological feature detection, and then editing and enriching the resulting datasets in the GeoPACHA webapp through our international network of regional experts and their diverse student teams. This approach will dramatically reduce the need for surveyors to scan grid squares with no visible loci, while placing people in the workflow where they can best contribute, as expert observers and analysts.

In summary, the team-based, problem-focused systematic imagery survey enabled by GeoPACHA has significantly broadened the frame for archaeological knowledge production in the central Andes. It has revealed continuous distributional vistas of settlement and land-use at scales that would otherwise be impossible. It has also opened up new questions and modes of questioning. We see encouraging trends for further scaling up our analyses through continued international collaboration—and, increasingly, through AI-assisted approaches, which will filter out featureless areas; enable surveyors to focus on potential loci; and identify, classify and register other observational data. Such an approach will not only provide even larger scale datasets, but also potentially reduce compromises between scale and data granularity, as surveyor time can be dedicated to making archaeological observations rather than reviewing featureless space. Yet such compromises will always exist. Imagery survey provides an extremely promising path forward for addressing some of archaeology's scalar challenges, both as a field of study in itself and as a complement to field archaeology, but it will always offer partial visions of archaeological landscapes that complement more detailed, field-based research. It is an additional layer of archaeological data that can serve as a high-level meshwork of distributional knowledge about past peoples and places.

Acknowledgements

We express our deepest appreciation to the many members of the GeoPACHA team for their many hours pursuing the development and execution of this project during especially challenging times. Our initial development efforts benefitted from consultations with James Artz, Jason Herrmann, Veronica Ikeshoji-Orlati and Rachel Opitz, and the technical expertise of Thanos Delas, Alex Drakos and John Wilson. While we are most grateful to all of our collaborators, any errors in this essay are solely ours.

Funding statement

Implementation-level funding for GeoPACHA was provided by an American Council of Learned Societies Digital Extension Grant (Steven A. Wernke, PI; Parker VanValkenburgh,

https://doi.org/10.15184/aqy.2023.177 Published online by Cambridge University Press

co-PI). Graduate student funding and machine learning model development were supported by NSF Grant Award 2106717 (Wernke, PI) and NSF Grant Award 2106766 (VanValkenburgh, PI). Initial development of GeoPACHA was supported by a National Endowment for the Humanities Level II Digital Humanities Startup Grant (Grant HD-229071-15, Wernke, PI), and a Center for Advanced Spatial Technology (CAST) Spatial Archaeometry Research Collaborations (SPARC) grant (Wernke and VanValkenburgh, co-PIs).

Data availability statement

The authors confirm that the data from this study are available from the corresponding author upon reasonable request. Data from the constituent survey projects will be made available to registered users via Open Context.

References

ADAMS, R.M. 1981. *Heartland of cities: surveys of ancient settlement and land use on the central floodplain of the Euphrates*. Chicago (IL): University of Chicago Press.

ALCOCK, S. & J. CHERRY. 2016. *Side-by-side survey: comparative regional studies in the Mediterranean world*. Oxford: Oxbow.

ALLEEN-WILLEMS, R. 2012. Designing the digital archaeological record: collecting, preserving, and sharing archaeological information. Unpublished PhD dissertation, Northern Arizona University.

ALTSCHUL, J.H. *et al.* 2018. Fostering collaborative synthetic research in archaeology. *Advances in Archaeological Practice* 6: 19–29. https://doi.org/10.1017/aap.2017.31

ANDERSON, D.G. *et al.* 2017. Sea-level rise and archaeological site destruction: an example from the southeastern United States using DINAA (Digital Index of North American Archaeology). *PLoS ONE* 12: e0188142. https://doi.org/10.1371/journal.pone.0188142

ARKUSH, E., L.E. KOHUT, R. HOUSSE, R.D. SMITH & S.A. WERNKE. 2023. A new view of hillforts in the Andes: expanding coverage with systematic imagery survey. *Antiquity*. Published online December 2023. https://doi.org/10.15184/aqy.2023.178

ATICI, L., S.W. KANSA, J. LEV-TOV & E.C. KANSA. 2013. Other people's data: a demonstration of the imperative of publishing primary data. *Journal of Archaeological Method and Theory* 20: 663–81. https://doi.org/10.1007/s10816-012-9132-9

BANNING, E.B. 2002. *Archaeological survey*. New York: Kluwer Academic/Plenum. https://doi.org/10.1007/978-1-4615-0769-7

BARKER, G., D. GILBERTSON, B. JONES & D. MATTINGLY. 1996. *Farming the desert: the UNESCO Libyan valleys archaeological survey II: site gazetteer and pottery*. New York: UNESCO, Society for Libyan Studies.

BAUER, B.S. & R.A. COVEY. 2002. Processes of state formation in the Inca heartland (Cuzco, Peru). *American Anthropologist* 104: 846–64. https://doi.org/10.1525/aa.2002.104.3.846

BEWLEY, R. *et al.* 2016. Endangered archaeology in the Middle East and North Africa: introducing the EAMENA project, in S. Campana, R. Scopigno, G. Carpentiero & M. Cirillo (ed.) *CAA2015. Keep the revolution going: proceedings of the 43rd annual conference on computer applications and quantitative methods in archaeology*, vol. 1: 919–32. Oxford: Archaeopress.

BICKLER, S.H. 2021. Machine learning arrives in archaeology. *Advances in Archaeological Practice* 9: 186–91. https://doi.org/10.1017/aap.2021.6

BLANTON, R., S. KOWALEWSKI, G. FEINMAN & J. APPEL. 1981. *Ancient Mesoamerica: a comparison of change in three regions*. Cambridge: Cambridge University Press.

BLANTON, R., G.M. FEINMAN, S.A. KOWALEWSKI & L.M. NICHOLAS. 1999. *Ancient Oaxaca*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511607844

CAO, B. *et al.* 2021. A 30m terrace mapping in China using Landsat 8 imagery and digital elevation model based on the Google Earth Engine. *Earth System Science Data* 13: 2437–56. https://doi.org/10.5194/essd-13-2437-2021

Casana, J. 2014. Regional-scale archaeological remote sensing in the age of big data: automated site discovery vs. brute force methods. *Advances in Archaeological Practice* 2: 222–33. https://doi.org/10.7183/2326-3768.2.3.222

– 2015. Satellite imagery-based analysis of archaeological looting in Syria. *Near Eastern Archaeology* 78: 142–52. https://doi.org/10.5615/neareastarch.78.3.0142

– 2020. Global-scale archaeological prospection using CORONA satellite imagery: automated, crowd-sourced, and expert-led approaches. *Journal of Field Archaeology* 45: S89–100. https://doi.org/10.1080/00934690.2020.1713285

Casana, J. & J. Cothren. 2013. The CORONA Atlas project: orthorectification of CORONA satellite imagery and regional-scale archaeological exploration in the Near East, in D.C. Comer & M.J. Harrower (ed.) *Mapping archaeological landscapes from space*: 33–43. New York: Springer. https://doi.org/10.1007/978-1-4614-6074-9_4

Casana, J. & M. Panahipour. 2014. Satellite-based monitoring of looting and damage to archaeological sites in Syria. *Journal of Eastern Mediterranean Archaeology & Heritage Studies* 2: 128–51. https://doi.org/10.5325/jeasmedarcherstu.2.2.0128

Caucasus Heritage Watch. 2022. The Armenian Cultural Heritage of Nakhchivan/Nakhichevan. Electronic document, https://caucasusheritage.cornell.edu/index.php/historic_research_details/3 (accessed October 2022).

Cherry, J.F. 2003. Archaeology beyond the site: regional survey and its future, in J.K. Papadopoulos & R.M. Leventhal (ed.) *Theory and practice in Mediterranean archaeology: old world and new world perspectives*: 137–59. Los Angeles: Cotsen Institute of Archaeology. https://doi.org/10.2307/j.ctvdjrrd6.14

Contreras, D.A. & J. Meadows. 2014. Summed radiocarbon calibrations as a population proxy: a critical evaluation using a realistic simulation approach. *Journal of Archaeological Science* 52: 591–608. https://doi.org/10.1016/j.jas.2014.05.030

Daniels, C. 1970. *The Garamantes of southern Libya*. London: Oleander.

Drennan, R.D., C.A. Berrey & C.E. Peterson. 2015. *Regional settlement demography in archaeology*. New York: Eliot Werner. https://doi.org/10.2307/j.ctvqc6hgm

Dunnell, R.C. 1992. The notion site, in J. Rossignol & LuAnn Wandsnider (ed.) *Space, time and archaeological landscapes*: 21–41. https://doi.org/10.1007/978-1-4899-2450-6

GlobalXplorer 2018. GlobalXplorer completes its first expedition: what the crowd found in Peru. *Medium*. Available at: https://medium.com/@globalxplorer/globalxplorer-completes-its-first-expedition-what-the-crowd-found-in-peru-7897ed78ce05. (accessed October 2018)

Hegmon, M. 2016. Archaeology of the human experience: an introduction. *Archeological Papers of the American Anthropological Association* 27: 7–21. https://doi.org/10.1111/apaa.12071

Johnson, G.A. 1977. Aspects of regional analysis in archaeology. *Annual Review of Anthropology* 6: 479–508. https://doi.org/10.1146/annurev.an.06.100177.002403

Kansa, E. 2012. Openness and archaeology's information ecosystem. *World Archaeology* 44: 498–520. https://doi.org/10.1080/00438243.2012.737575

Kansa, E.C., S.W. Kansa, J.J. Wells, S.J. Yerka, K.N. Myers, R.C. DeMuth, T.G. Bissett & D.G. Anderson. 2018. The Digital Index of North American Archaeology: networking government data to navigate an uncertain future for the past. *Antiquity* 92: 490–506. https://doi.org/10.15184/aqy.2018.32

Kansa, S.W. & E.C. Kansa. 2007. Open content in open context. *Educational Technology* Nov–Dec: 26–31.

Kansa, S.W., E.C. Kansa & J.M. Schultz. 2007. An open context for "Near Eastern archaeology". *Near Eastern Archaeology* 70: 188–94. https://doi.org/10.1086/NEA20361331

Kantner, J. 2008. The archaeology of regions: from discrete analytical toolkit to ubiquitous spatial perspective. *Journal of Archaeological Research* 16: 37–81. https://doi.org/10.1007/s10814-007-9017-8

Kintigh, K.W. & J.H. Altschul. 2010. Sustaining the digital archaeological record. *Heritage*

https://doi.org/10.15184/aqy.2023.177 Published online by Cambridge University Press

Management 3: 264–74. https://doi.org/10.1179/hma.2010.3.2.264

LIN, A.Y.-M., A. HUYNH, G. LANCKRIET & L. BARRINGTON. 2014. Crowdsourcing the unknown: the satellite search for Genghis Khan. *PLoS ONE* 9: e114046. https://doi.org/10.1371/journal.pone.0114046

MARCONE, G., G. HUERTAS, J. ZIMMER-DAUPHINEE, P. VANVALKENBURGH, J. MOAT & S.A. WERNKE. 2023. Late pre-Hispanic fog oasis settlements and long-term human occupation on the Peruvian central coast from satellite imagery. *Antiquity*. Published online December 2023. https://doi.org/10.15184/aqy.2023.179

MCCOY, M.D. 2020. The site problem: a critical review of the site concept in archaeology in the digital age. *Journal of Field Archaeology* 45: S18–26. https://doi.org/10.1080/00934690.2020.1713283

MCMANAMON, F.P., K.W. KINTIGH, L.A. ELLISON & A. BRIN. 2017. tDAR: a cultural heritage archive for twenty-first-century public outreach, research, and resource management. *Advances in Archaeological Practice* 5: 238–49. https://doi.org/10.1017/aap.2017.18

MENZE, B.H. & J.A. UR. 2012. Mapping patterns of long-term settlement in Northern Mesopotamia at a large scale. *Proceedings of the National Academy of Sciences USA* 109: E778–87. https://doi.org/10.1073/pnas.1115472109

MILLS, B.J. & W.H. WALKER (ed.) 2008. *Memory work: archaeologies of material practices*. Santa Fe (NM): School for Advanced Research.

MORETTI, F. 2013. *Distant reading*. Brooklyn (NY): Verso.

PARCAK, S. 2009. *Satellite remote sensing for archaeology*. New York: Routledge. https://doi.org/10.4324/9780203881460

– 2019. *Archaeology from space: how the future shapes our past*. New York: Henry Holt and Co.

PERREAULT, C. 2019. *The quality of the archaeological record*. Chicago (IL): University of Chicago Press. https://doi.org/10.7208/chicago/9780226631011.001.0001

RODDICK, A.P. & A.B. STAHL (ed.) 2016. *Knowledge in motion: constellations of learning across time and place*. Tucson: University of Arizona Press.

ROSS, S., A. SOBOTKOVA, B. BALLSON-STANTON & P. CROOK. 2013. Creating eresearch tools for archaeologists: the federated archaeological information management systems project. *Australian Archaeology* 77: 107–19. https://doi.org/10.1080/03122417.2013.11681983

SANDERS, W.T. 1970. *The Teotihuacan Valley project final report*, vol. 1. State College: Pennsylvania State University.

SANDERS, W.T., J.R. PARSONS & R.S. SANTLEY. 1979. *The basin of Mexico: ecological processes in the evolution of a civilization*. New York: Academic Press.

SOROUSH, M., A. MEHRTASH, E. KHAZRAEE & J.A. UR. 2020. Deep learning in archaeological remote sensing: automated qanat detection in the Kurdistan region of Iraq. *Remote Sensing* 12: 500. https://doi.org/10.3390/rs12030500

SPENCE MORROW, G., P. VANVALKENBURGH, C. WAI, & S.A. WERNKE. 2023. Augmenting field data with archaeological imagery survey: mapping hilltop fortifications on the north coast of Peru. *Antiquity*. Published online December 2023. https://doi.org/10.15184/aqy.2023.176

SPIELMANN, K.A. & K.W. KINTIGH. 2011. The Digital Archaeological Record: the potentials of archaeozoological data integration through tDAR. *SAA Archaeological Record* 11: 22–25.

STANISH, C. 1987. Agroengineering dynamics of post-Tiwanaku settlements in the Otora Valley, Peru, in W.M. Denevan, K. Mathewson & G. Knapp (ed.) *Pre-Hispanic agricultural fields in the Andean region* (British Archaeological Reports International series 359): 337–64. Oxford: Archaeopress.

SUPERNANT, K., J.E. BAXTER, N. LYONS & S. ATALAY. 2020. *Archaeologies of the heart*. New York: Springer. https://doi.org/10.1007/978-3-030-36350-5

UR, J.A. 2006. Google Earth and archaeology. *SAA Archaeological Record* 6: 35–38.

WHITLOCK B., P. VANVALKENBURGH & S.A. WERNKE. 2023. Managing pastoral landscapes: remote survey of herding infrastructure in Huancavelica, Peru. *Antiquity*. Published online December 2023. https://doi.org/10.15184/aqy.2023.174

WELLS, J.J., E.C. KANSA, S.W. KANSA, S.J. YERKA, D.G. ANDERSON, T.G. BISSETT, K.N. MYERS & R. CARL DEMUTH. 2014. Web-based discovery and integration of archaeological historic properties inventory data: the Digital Index of North American Archaeology (DINAA). *Literary and Linguistic Computing* 29: 349–60. https://doi.org/10.1093/llc/fqu028

Zimmer-Dauphinee, J., P. VanValkenburgh & S.A. Wernke. 2023. Eyes of the machine: AI-assisted satellite archaeological survey in the Andes. *Antiquity*. Published online December 2023. https://doi.org/10.15184/aqy.2023.175