The Emerging Open World

Martin Hall

into which Kansa and Whitcher Kansa see their puzzle pieces fitting, however, assumes an access to and a level of funding that I doubt that humanists typically experience. So, how might their model of funding and publication work in other venues in the humanities? If humanities scholars could count on even a fraction of $800,000 to publish their research projects, it would be magical. Projects could be created for open access from the outset. Absent funding this would not work so well. Here's where the humanities road gets a little rocky. Public funding for literature, American studies, history, philosophy, art history, and the like is hard to come by. For instance, the National Endowment for the Humanities, which, in part, supported Kenan Tepe, funds the humanities. But differences in the level of funding required (see p. 95 of Kansa and Whitcher Kansa)—and amount available—cannot be overlooked. Moreover, NEH is not known for funding scholarly publishing.[2] Another governmental funding agency, the National Endowment for the Arts, operates similarly. Neither of these adequately supports scholarly publishing in the humanities.[3] Kenan Tepe required funding over a period of many years; understandably, the excavation also sought resources in a variety of places, both public and private.[4] Moreover, although the costs or the funding needed in the disciplines like those listed above might not be close to the needs in archaeological excavations, one can assume the cost of publishing humanities research would not be less.[5] Let's say costs for a humanities monograph also come in around $10,000–$15,000 (that does not count the overhead costs, which would also amount to another $10,000–$15,000). Kenan Tepe publication costs represent a "small fraction" (1.87%) of the total amount of funding received for the excavation. Open access seems an easy choice. Troublingly, the typical humanities scholar (and publisher) rarely receives even a "small fraction" of an $800,000 tax-payer funded grant to fund publication. Thus, open access publishing in the humanities, without adequate funding, will remain difficult.

Notes

1. Fundamentally it is disingenuous for open access advocates to frame the problem solely as a moral issue, having to do with dissemination or public good. One can be a proponent of free access and still realize that the costs are real and must be accounted for. Scholarly associations and organizations, especially in the arts and humanities, have traditionally supported dissemination and access. They continue to do so. The dearth of taxpayer-supported funding for arts and humanities publishing makes it difficult to overcome financial challenges intrinsic to open access.
2. A search of the NEH database for funded projects under the category scholarly publications from 2000–2012, for example, yielded no results. Scholarly publications may be funded under other grants, but this points to the basic problem.
3. The differences in governmental funding levels for the sciences v. the arts and humanities are telling. Compare below the 2013 requests: NIH: FY 2013, $31B (http://officeofbudget. od.nih.gov/); NSF: $7.4B; (www.nsf.gov/.../budget/fy2013/ pdf/EntireDocument_fy2013); NEA: $154M (http://www.nea. gov/news/news12/Budget.html); NEH: $154M (www.neh.gov/ files/neh_request_fy2013.pdf).
4. See for example, Bradley J. Parker, Catherine P. Foster et al., "The Upper Tigris Archaeological Research Project (UTARP): A Preliminary Report from the 2007 and 2008 Field Seasons at Kenan Tepe," *Anatolica* 35 (2009): 85–152.
5. See Mary Waltham, "The Future of Scholarly Journals Publishing among Social Science and Humanities Associations: Report on a Study Funded by a Planning Grant from the Andrew W. Mellon Foundation" (Princeton, NJ, 18 February 2009).

RESPONSE

The Emerging Open World

MARTIN HALL

Vice-Chancellor, University of Salford
The Old Fire Station, The Crescent, Salford M5 4WT, United Kingdom, martin.hall@salford.ac.uk

An overarching aspect of our contemporary world is this. The transition in the use, and abuse, of digital information is so comprehensive and profound that it is largely taken as commonplace. This affects archaeology as it does every other field of inquiry. We take for granted that we can move a terabyte or two of data around with us and that we can pull down information onto cheap devices that know where we are, almost anywhere in the world. We also expect to be able to find anything we want, from anywhere in the world, on the Internet and we become impatient if access takes more than a minute. And yet, for any adult alive today, these transitions have taken place

within her lifetime. A person who was aged ten in 1990 was growing into a world with nothing like the digital resources we take for granted today. None of us would accept a return to the standards of digital information that we found ground-breaking just ten years ago. And within five years we will find unacceptable the standards of data access that we receive today.

Archaeology has, of course, always been a data-heavy discipline and, because of this, archaeologists were early adopters of digital technologies. In the early 1970s we were laboriously punching cards for overnight batch runs. Ten years later we were hauling early personal computers into the field. By the early 1990s, reasonable quality images of our artifacts were widely available online. Ten years ago we were publishing extensively on the Web. The issues around digital data and its dissemination are hardly new. Neither are questions about data publication. Most jurisdictions require that excavation reports and data are lodged with statutory authorities and that, following analyses, assemblages are housed in recognized museums or equivalent institutions, where they should be curated to acceptable standards. Issues of rights of access to these data sets—whether field notes and reports or the objects themselves—have long been argued over. In this sense, the recent revolution in the storage capacity of digital data and the ease, speed, and volume of online dissemination is less a paradigm shift than a step change in order of magnitude.

What is new though—and which makes Kansa and Whitcher Kansa's paper and its discussion timely—is open data. In its essence, the advent of open data is a return to Thomas Jefferson's dictum that the essential property of knowledge is that it can be used and consumed without destroying the original (unlike a can of beans or a book). And digital data has no original in that every one of an infinite number of perfect copies is a simulacrum. Recent technological advances have made it, in practice, impossible to contain or reserve exclusive use of digital data, as music companies and film distributors have discovered. The open data movement recognizes this, and seeks new protocols for recognizing and acknowledging rights (including monetary rights), and verifying authenticity and quality. Kansa and Whitcher Kansa are right to claim

that, for a data-intense discipline such as archaeology, the implications are profound.

These developments in the management and distribution of digital data cannot be separated from equally profound developments in scholarly publishing. As we all know, conventional scholarly publishing is based on subscription journals, formerly printed and posted, but now much more often available to subscribers online. Many academics have taken the shift from paper copies to online access for granted, maybe even assuming that online editions are "free." This is because of disintermediation or, more prosaically, the increasing invisibility of the librarian. In reality, you can get the *Journal for Eastern Mediterranean Archaeology and Heritage Studies* online, without paying, because your university has bought a subscription and the journal recognizes your Internet address as legitimate. The problem is that, overall, online journal prices have been escalating, year-on-year, well above the general rate of inflation. Because this is not sustainable, the scholarly world is moving to open access publishing, through both "green" repositories (in which the final draft is deposited by the author before surrendering copyright to a publisher) or via the "gold" route (in which the full cost of publication is paid for up-front, allowing the published version to be available to anyone online, without payments or other constraints).

Over the next few years, green and gold will converge, and all publication costs will be paid up front. This will be important because of the huge advances that have been made in data mining. Data mining is done by automated digital robots that move rapidly and constantly through open data fields, answering questions. This is already transforming research fields such as epidemiology, finding new patterns that point to the causality of medical conditions. Data mining depends on a barrier-free internet—online robots cannot work effectively if they keep encountering demands for a $35 fee to read a journal article. Effective data mining has obvious implications for archaeology, allowing new generations of metadata analysis ranging from trawls for DNA patterns in all known excavated skeletal material to trace analysis of ceramic glazes in collections held across the world's museums.

Taking Kansa and Whitcher Kansa's discussion further, then, is to note that the issues that they correctly raise around data publication cannot be separated from

publication more generally. We can anticipate a world in which all scholarly publications are fully open access, and in which each publication is expected to include hyperlinks to the open data sets on which it is based. We can also expect many of these services to be facilitated by business intermediaries, whether non-profit or for profit, raising a host of secondary, and ethical, issues both for archaeologists and for the institutions that hire them.

RESPONSE

Sharing Data is Hard!—But Worth It

BENJAMIN W. PORTER

University of California, Berkeley

I developed one of those finish-the-sentence chants to teach my young son to share with other kids. "Sharing is . . . ," I say, raising my voice to prompt him, and he shouts ". . . good!" Of course, we all know that sharing is . . . hard! But willingly or not, sharing is the new ethic for a new century, one that archaeologists, regardless of age, must adapt to for so many reasons described in Kansa and Whitcher Kansa's article. It is like preschool all over again for archaeologists, and our authors are those patient teachers urging us to share with each other. This time it is our most precious toys we must lend: our data.

But if data publication is to have a promising future, then advocates like Kansa and Whitcher Kansa must reflect deeply on the cultures of academic sharing. Sharing is cultural, after all, and like all culture, it is mostly learned, and hardly innate. The willingness to distribute the materials that make up the backstage of final publications—the gray literature, the spreadsheets, the images, the really-good-but-rejected grant application— is a habit that emerges through specific disciplinary formations, practices that are learned from mentors and peers early in one's career.

So I believe our authors are still trying to decipher archaeology's far from homogeneous attitudes toward sharing. Case in point is the way they mistakenly typecast their audience as academics pressured to manufacture only traditional print-run publications in the highest ranked journals. Scholars supposedly have no time for projects like data sharing that bring little credit from university promotion committees. This is partly true, of course. But this type-cast scholar, a highly rational hamster, wheel-peddling her way to tenure, is somewhat of a straw man. Academics—like this author, who is writing an un-refereed response in the first issue of a fledgling journal—make contributions to knowledge that are rarely acknowledged in their promotion cases.

In order to get at the crux of this issue, one should not ask why archaeologists don't share, but rather, why (and when) do they share? Any pretense of altruism should be discarded immediately; studies on sharing, cooperation, and collaboration demonstrate that when people give, they implicitly expect reciprocity, maybe not from the immediate receiver, but from some other source at a later time. Data is therefore a type of capital (*sensu* Bourdieu) that is displayed and exchanged in ways that boost the reputation of scholars and their projects. If one has BIG data sets, their display leaves admirers awe-struck and competitors envious. Or the discovery of a singularized datum that is reported in the media—an inscription or a monumental sculpture—boosts a scholar's reputation in the public eye. Archaeologists obviously surrender their capital when it is strategically beneficial for them or their projects. But the calculated timing of such events is potentially revealing for data-sharing advocates. More research is needed to identify the co-occurring circumstances and pressures that give rise to these moments of sharing.

A few of my early mentors warned me not to share data before "final" publication. I remember us scrambling when a supposed adversary was visiting. Lock the door, hide the keys, deflect questions. I was told that if we hid the materials until the last possible minute, we could then retain control over its interpretation. That data is still not published (the adversary's is). But now I do share, for so many reasons, beyond the fact that it earns me capital. I believe the authors are correct that sharing data in open-access platforms is a public good, especially for those working at public institutions or with public funding. I also share for practical reasons: my office in Berkeley sits above the Hayward Fault and when the Big One happens, all of my notes, books, and data may be buried under nine floors of office debris. It helps me sleep easier at night knowing