

# Making the Most of Cyberinfrastructure

Dean R. Snow  
The Pennsylvania State University

## Abstract

Problems in cyberinfrastructure development fall into strategic, tactical, and technical categories. I argue that we tend to focus on the last and neglect the other two, to the potential detriment of archaeology. Disciplines vary in what are considered data, the sizes of data sets, the ease of data acquisition, standards of confidentiality, when and how data should be made public, and opinions regarding what should be preserved over the long term. Current efforts to foster cyberinfrastructure in the social and behavioral sciences are dominated by social scientists who tend to work with large survey data sets on a narrow range of problems. While there is considerable variation in standards and objectives within anthropology alone, anthropological standards and practices fall largely outside the realm of the dominant social sciences. Standards and practices in archaeology are at particular risk of being marginalized as science moves forward in this critically important endeavor. This paper discusses ways in which archaeologists can establish links to other disciplines having similar strategic and tactical goals so that archaeology is not left behind in the larger cyberinfrastructure effort.

## Introduction

I argue that we are well advised to divide problems in cyberinfrastructure development into strategic, tactical, and technical categories. My reasoning is that we tend to focus on technical problems and to bypass important strategic and tactical problems in the process. By dividing problems into these three categories we at least ensure that we will recognize that the first two problem areas exist as significant constraints on solutions found to problems in the technical category. I have already outlined this approach elsewhere, and there is a good chance that it will have been published by the time of the 2010 SAA meeting in St. Louis (Snow 2010). Consequently I will only outline the larger argument here.

## Facilitate the Unpredictable

They called it a “horseless carriage” because they could not initially imagine that the automobile would be anything more than that. One early take on television was that the new medium was just like radio only you could see who was talking. People of my generation might be forgiven if they adopt a similarly myopic view of archaeoinformatics, but we should all understand that to the generation of archaeologists currently rising it will not simply be a matter of digitizing and automating what we have traditionally done in the way of archiving, data sharing, and synthesis. We are in a new age of combinatorial innovation that will prove in retrospect to be profoundly new and different (Varian 2009). This is truly an emergent phenomenon and it is unlikely that we will be able to predict it accurately in all or even some of its eventual dimensions. There is no longer a need to argue in favor of cyberinfrastructure; that has been established. What we need to worry about now is fostering its development in ways that will not be constrained by our current limited perspectives.

Some specific points are important to keep in mind:

1. Information was once scarce and expensive but relatively reliable.
2. Information is now abundant, cheap, but less reliable.
3. There is a poverty of attention in a world awash with information.
4. Facts do not speak for themselves, never have and never will.
5. Most users are so overwhelmed by the information that is already available that they give up trying to use it.
6. Specific cybertools will be developed with or without our strategic vision, and some of them will spread virally.

### **Facilitating the Right Strategic Approach**

Archaeology shares some important features with the other historical sciences. We are data poor compared to many other sciences, and acquiring new data is costly. Thus we are dependent upon archived data, which was typically gathered for reasons not entirely congruent with our research designs. Many if not all of our conclusions are inferential, and to be convincing they must be based on testable theory derived from modern observation of real world phenomena. Because of this historical scientists can never verify observations in the same way the other sciences can, and this has led at least some people to opine that historical science is not science at all, or at least not worth their prolonged attention.

Archaeology is typically considered to be a part of anthropology, which in turn is typically assigned to the social sciences. But social scientists who conduct survey research, which is most of them, often hold legacy data in low regard. Archaeology, with its paltry data and its dependence upon legacy data, is viewed as quaint and completely lacking in rigor by many survey scientists. Although we and they are generally regarded as colleagues in the social sciences, their methods differ from ours and are likely to garner them the lion's share of research support going forward. Our natural allies are in geology, paleontology, paleoecology, and other historical sciences. It is with them that we should partner and it is from them that we are most likely to get good new ideas.

The best strategy for archaeology is to create systems and services that will:

1. Preserve key digital files in standard formats
  - a. Databases
  - b. Documents
  - c. Images
  - d. GIS layers
2. Make gray literature accessible
3. Promote interoperability
4. Archive both positive and negative findings
5. Foster the development and incorporation of "viral" tools

Archaeologists must do their best to coordinate these activities because funding will be severely constrained and we will not have the luxury of trying out competing systems. This strategy risks the inadvertent creation of a comprehensive approach that will inhibit innovation and evolution. What we need is openness and flexibility. What we do *not* need are closed systems or standardization mandates. As part of our overall strategy we should encourage open source software, translation tools, attractive sharing tools, and a generally distributed system (as opposed to a centralized one).

### **Tactics that Facilitate**

Tactically speaking, we have several new and not so new services at hand. Archaeological Data Services (ADS) already exists in the UK. The Alexandria Archive has been in operation for a few years. The National Archaeological Database has been around longer. Now we have Digital Antiquity almost ready to launch and a team in place to make it work.

But not all of these have worked out or will work out as well as hoped. Some services not mentioned have already disappeared. Some still persisting are underutilized or out of date. It is clear that we cannot simply take the position that we can “build it and they will come.” Perfectly good systems will languish unused unless they are designed and rolled out properly.

Why have some good systems failed? There are probably many reasons but here are a few likely ones:

1. Some systems depend upon voluntary contributions that are difficult to submit and for which there is no clear benefit to the donor.
2. The pool of potential users is large and anonymous rather than comprised of a small set of known potential users with whom one is already collaborating.
3. Some systems lack tools for easy conversion of data to make submitted databases interoperable.
4. Systems that are presented as centralized repositories are less attractive than systems that are presented as means to link distributed repositories.
5. Researchers want to retain control of their data, images, and documentation as continuously editable files.
6. Younger scholars tend to protect their data as their still limited stock in trade.
7. No matter how old a researcher is (s)he does not think (s)he is done yet.
8. There is a widespread perception that publication earns credit but sharing often does not.

Because of these factors, we must adopt tactics that will attract, not threaten, the sensibilities of researchers whose records we would like to see archived and shared. I argue that we should adopt some specific tactics to ensure adoption of the system(s) we create.

1. Accommodate repositories that are distributed and likely to resist centralization.
2. Provide assistance to database managers who seek to protect confidential information such as site locations while making other associated data available.
3. Encourage early adoption by small groups of cooperating researchers who are already collaborating by promoting the system as a means to make what they are already doing easier.
4. Provide for hierarchical access. Any new system should be useful and attractive to professional archaeologists, but nonconfidential parts of it should also be open to the general public.
5. Promote the view that it is the ethical responsibility of professional archaeologists to provide up-to-date syntheses of data that will otherwise be overwhelming for nonprofessionals who access it.
6. Professional mediation should also supplement automatic conversions so that, for example, users can derive GPS coordinates from files originally encoded using the township and range system.

7. Professional journals should be encouraged to credit access to electronic resources in ways similar to standard publication citation.
8. Encourage funding agency requirements for project data preservation and access as a prerequisite for new funding.
9. Work with agency, commercial, and academic repositories on the preservation of electronic legacy data.
10. Make training in preservation and archiving as important as training in acquisition, analysis, and publication. Graduate training should emphasize all the following steps.
  - a. Research design
  - b. Data acquisition
  - c. Data Analysis
  - d. Publication
  - e. Digitization
  - f. Deposition
    - i. Preservation
    - ii. Archiving
    - iii. Access

### **Technical Facilitation**

This is where discussion usually starts, but I argue that it is the least important aspect of the larger question of where archaeoinformatics is or should be going. We already have an array of technical tools available to us. I have already discussed ArchSeer, a specialized archaeological search engine, in other publications (Snow, et al. 2006). I might also mention tools like Google Earth, JSTOR, and Google Books. More comprehensive services like ADS and Digital Antiquity are likely to incorporate tools like these as they develop, and such tools will make the services increasingly attractive to users. The technical stuff will almost take care of itself, and I will provide a couple examples to illustrate what I mean.

*Example 1.* Some technical tools are likely to go viral on their own. One example is 1<sup>st</sup> Photo Marker, viewable at <http://promoappcodes.com/apps/1st-photo-marker>. This is an iPhone app that began as a simple app for taking photos with a few attached data fields visible. I contacted the developer and it did not take much persuading to get him to add some additional features. One can now use the app to take a photo that has attached to it and selectively visible on it:

1. Title
2. Subject
3. Optional additional user-defined fields
4. ID No.
5. Date
6. Time
7. GPS coordinates
8. Distance to a selectable home base
9. Aspect
10. Altitude

This is a terrific tool for anyone gathering field data, but there is more. The app has a built-in feature that allows the user to instantly send the image with all of its associated data to a stored email address chosen by the user. It is not difficult to imagine a SHPO office receiving many

such images from employees, professional archaeologists, or even the general public. The question is not how we go about implementing such an app; it is available to anyone for only 99 cents and is likely to be purchased by any archaeologist with an iPhone as soon as (s)he knows about it. The real question is how to we manage and share the mass of data such an app is likely to generate? This is the part that will not take care of itself.

*Example 2.* Another example is EditMe, a wiki hosting site at <http://www.editme.com/>. In my judgment it would be a good idea for the SAA to host a wiki for professional mediation of archaeological data. An expensive NSF scheme to make biological data available to secondary school teachers and other people in the general public has largely failed because the potential users are overwhelmed by the mass of data that a simple search produces. One discouraged middle school teacher said “I’d rather have a nice-sized catalog of peer-reviewed material that promotes active learning than a vast amount of stuff that hasn’t been vetted” (Mervis 2009). Teachers have immediate needs and casual users have short attention spans. Neither can make effective use of unprocessed information. Most will turn to Wikipedia, where synthetic entries have not necessarily been produced by qualified archaeologists. Traditional specialized encyclopedias used to solve this problem, but they are expensive, hard to find and access, quickly out of date, and facing extinction. A specialized archaeological wiki would solve this problem. We could maintain such a wiki with entries written at two levels. For qualified professionals, and available only through the SAA member login, the wiki could provide up-to-date synthetic articles with all the relevant information. Within each entry there could be a box to contain confidential information such as site location. These boxes could be excluded from a publically accessible version of the wiki. Who among us would shrink from providing short wiki entries for the sites on which we have already expended no small amount of blood, sweat, and tears? What middle school teacher would not appreciate having access to brief reliable information and clear cultural context on an interesting site? How do we best inspire ourselves to produce this new kind of SAA publication?

These are just two tools that I think we can and should be incorporating into our strategic and tactical thinking as we try to stay ahead of cyberinfrastructure developments in archaeology. It will be a mistake to let events unfold on their own. The avoidance of expensive bad ideas should inspire us to think and act carefully. I am pleased that this electronic symposium is taking a lead.

### References Cited

- Mervis, J.  
2009 NSF Rethinks its Digital Library. *Science* 323:54-58.
- Snow, D. R.  
2010 Making Legacy Literature and Data Accessible in Archaeology. In *Proceedings of the CAA 2009*, edited by B. Frischer, pp. 350-355. Computer Applications in Archaeology.
- Snow, D. R., M. Gahegan, C. L. Giles, K. G. Hirth, G. R. Milner, P. Mitra and J. Z. Wang  
2006 Cybertools and Archaeology. *Science* 311:958-959.
- Varian, H.  
2009 Hal Varian on how the Web challenges managers. In *the McKinsey Quarterly*. McKinsey & Company.