

# CYBER INFRASTRUCTURE FOR THE HUMANITIES AND SOCIAL SCIENCES

**John Unsworth<sup>1</sup>**

In January 2003, a blue ribbon panel appointed by the National Science Foundation and led by Dan Atkins, of the University of Michigan, completed a report called ‘Revolutionising Science and Engineering through Cyberinfrastructure’.<sup>2</sup> This report is a kind of provocation for the American Council of Learned Societies (ACLS) Commission on Cyberinfrastructure for the Humanities and Social Sciences, and there is a lot of other activity of this sort, going on right now—for example:

- Digital Archiving and the National Archives and Records Administration;<sup>3</sup>
- NSF ‘Post Digital Library Futures’ report;<sup>4</sup>
- NRC ‘Beyond Productivity’ report (2003);<sup>5</sup>
- The United Nations World Summit on the Information Society.<sup>6</sup>

---

<sup>1</sup> Chair, Commission on Cyberinfrastructure for the Humanities and Social Sciences, American Council of Learned Societies; Dean and Professor, Graduate School of Library and Information Services, University of Illinois at Urbana-Champaign.

This chapter was first presented as a paper at the annual meeting of the Research Libraries Group (Washington, DC, 26 April 2004).

<sup>2</sup> Daniel Atkins et al, *Revolutionizing Science and Engineering through Cyber-infrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure* (January 2003) <[http://www.communitytechnology.org/nsf\\_ci\\_report/](http://www.communitytechnology.org/nsf_ci_report/)>.

<sup>3</sup> Digital Archiving and the National Archives and Records Administration (2004) <[http://www7.nationalacademies.org/cstb/project\\_nara.html](http://www7.nationalacademies.org/cstb/project_nara.html)>.

<sup>4</sup> National Science Foundation, *Knowledge Lost in Information: Report of the NSF Workshop on Research Directions for Digital Libraries* (2003) <<http://books.nap.edu/books/0309088682/html/index.html>>.

<sup>5</sup> National Research Council of the National Academies, *Beyond Productivity: Information Technology, Innovation and Creativity* (2003) <<http://books.nap.edu/books/0309088682/html/index.html>>.

---

In a press release that followed the publication of the Atkins report, Peter Freeman, the assistant director of Computer and Information Sciences and Engineering (CISE), the NSF directorate that commissioned the report, said: ‘The path forward that this report envisions. . . truly has the potential to revolutionise all fields of research and education.’ Certainly, the report has had a significant impact on the rhetoric, and perhaps also on the priorities, not only of CISE, but also of other parts of NSF, and on other funding agencies concerned with information technology as it supports research.

So, what *is* cyberinfrastructure? Here is how the Atkins report addresses that question:

The term infrastructure has been used since the 1920s to refer collectively to the roads, power grids, telephone systems, bridges, rail lines, and similar public works that are required for an industrial economy to function. Although good infrastructure is often taken for granted and noticed only when it stops functioning, it is among the most complex and expensive thing that society creates. The newer term cyberinfrastructure refers to infrastructure based upon distributed computer, information and communication technology. If infrastructure is required for an industrial economy, then we could say that cyberinfrastructure is required for a knowledge economy.<sup>7</sup>

Cyberinfrastructure then is the infrastructure for a knowledge economy. And why should we care about it? Well, we all live, and will continue to live, in that knowledge economy, so we all have at least the same interest we would have in good roads and bridges, good telephone systems and power grids. And why should the humanities and social sciences care about it? Because we can make it a better infrastructure, if our perspectives, our training, and our expertise are included in its design and deployment. After all, science—whose goal is predictive certainty—only has half the picture. Uncertainty (or ambiguity, if you prefer) is the

---

<sup>6</sup> United Nations, *World Summit on the Information Society* (2003, 2005)  
<<http://www.itu.int/wsis/>>.

<sup>7</sup> Daniel Atkins et al, *Revolutionizing Science and Engineering through Cyber-infrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure* (January 2003) 5  
<[http://www.communitytechnology.org/nsf\\_ci\\_report/](http://www.communitytechnology.org/nsf_ci_report/)>.

other half, and the humanities and social sciences celebrate that, explore it, tolerate it, and *understand* it better than the sciences do. Or, at another level, if science and engineering are about what we can do, the humanities and social sciences are about what we should do. If we do not know what we can do, we do not know what choices to consider, but if we do not know what we should do, we do not know which choices to make. Cyberinfrastructure is no different, in that respect, from atomic energy, biotechnology, or any other challenge: it is not only a scientific challenge, with scientific outcomes: it is also a social and human challenge, with outcomes that the humanities and social sciences are best equipped to understand.

The ‘overarching finding’ of the Atkins report

... is that a new age has dawned in scientific and engineering research, pushed by continuing progress in computing, information, and communication technology, and pulled by the expanding complexity, scope, and scale of today’s challenges. The capacity of this technology has crossed thresholds that now make possible a comprehensive ‘cyberinfrastructure’ on which to build new types of scientific and engineering knowledge environments and organisations and to pursue research in new ways and with increased efficacy. Such environments and organisations, enabled by cyberinfrastructure, are increasingly required to address national and global priorities, such as understanding global climate change, protecting our natural environment, applying genomics-proteomics to human health, maintaining national security, mastering the world of nanotechnology, and predicting and protecting against natural and human disasters, as well as to address some of our most fundamental intellectual questions such as the formation of the universe and the fundamental character of matter.<sup>8</sup>

I agree with all of this, and I am certain that Dan Atkins, and many other scientists, would agree that along with all this new knowledge, these new certainties, will come new uncertainty, and new quandaries, that science

---

<sup>8</sup> Daniel Atkins et al, *Revolutionizing Science and Engineering through Cyber-infrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure* (January 2003) 31 <[http://www.communitytechnology.org/nsf\\_ci\\_report/](http://www.communitytechnology.org/nsf_ci_report/)>.

itself, by itself, will not be able to resolve. But if the humanities and social sciences want to have some influence in the process now underway to design our information technology environment over the next decade, then we need to articulate our needs and our potential contributions—and even more than that, we need to articulate the importance of the humanities and the social sciences, for the amelioration of the human condition. That is something we have not done very well since progress displaced enlightenment as our culture's highest value, and it is something that we still do not do very well.

For example, in the 'Summary of 2003 Fiscal Year Budget Request', the NEH argues for the humanities as follows:

In the 1965 legislation that established the National Endowment for the Humanities, the Congress of the United States declared that 'democracy demands wisdom and vision in its citizens' and posited that 'promoting progress in the humanities' was the surest route to such wisdom ... The National Endowment for the Humanities helps Americans develop wisdom and vision through the study and contemplation of the record of human thought. The study of history, literature, languages, philosophy and other humanities subjects help us not only to better understand our own nation, but other cultures as well.<sup>9</sup>

All well and good, and I believe also true, but unfortunately we in the United States seem to believe that wisdom is a great deal less expensive than knowledge, and knowledge—especially knowledge with practical consequences—is what we are willing to spend money on.

But the humanities and the social sciences have access to knowledge that does have enormous practical consequence, and the future will be better, or worse, depending on whether that knowledge is part of our 'knowledge society'. In order to be a meaningful part of that society, I would argue, the humanities and social sciences will need computational methods and they will need access to the kind of vast datasets that make computational methods both necessary and useful. Computational methods already have a place in the social sciences, and they have a foothold now in the study of literature, history, art, and other humanities

---

<sup>9</sup> National Science Foundation, *Summary of FY 2003 Budget Request to Congress* <[http://www.nsf.gov/bio/budget/bio\\_bdg03/bio\\_bgt03\\_toc.htm](http://www.nsf.gov/bio/budget/bio_bdg03/bio_bgt03_toc.htm)>.

disciplines. The humanities, in particular, has been without a galvanising methodology for a generation now, and it is being, and will be, revolutionised by information technology as profoundly as any of the sciences.

That revolution, though, has some preconditions: it requires a *motivating factor*, to move the disciplines toward new methods, and it requires the *means*—both intellectual and financial—to adopt, refine, and disseminate those methods for the rising generation of scholars.

The motivation, if it comes, will come in the form of very large datasets that can only be manipulated and interpreted with the aid of computers. We are getting there, already, with digital libraries, and if projects like DLF's distributed digital library come about, we will seem suddenly to have arrived. But even that scale is not quite what I have in mind. To arrive at terabytes or petabytes of humanities or social science data, we will have to effectively address two issues—intellectual property (the primary data-resource-constraint in the humanities) and privacy rights (the primary data-resource-constraint in the social sciences). In a sense, as John King pointed out to me, this is no different from the struggle, in computational science, with resource constraints on memory, bandwidth, or processing speed: unless we radically increase these resources, we radically limit the kinds of questions we can ask and answer. We generally think of intellectual property and privacy rights as legal issues, which of course they are, but short of sweeping legal remedies, which I do not expect and do not actually desire in either case, I think the solutions to these problems will be, to some significant extent, technical. That makes them, in effect, the primary 'cyberinfrastructure' research agenda for the humanities and social sciences. If we tackle that agenda successfully, the humanities will have access to the full recorded history of the 20<sup>th</sup> century—music, film, text and image, all in digital form; the social sciences will have access to the full record of societies, populations, individuals. There is tremendous danger of abuse, here—as there is with any other research that has profound practical consequences—but there is also a tremendous opportunity to learn, to understand, even to achieve some wisdom. Lest that emphasis on big datasets sound too much like 'rugged informationalism', let me emphasise that I agree with David Weinberger that 'what the world needs [is] people who know how to manage metadata, navigating the twisty darkness of the ambiguous world while

preserving the value of the unspoken.’ I would just add that if we only have access to metadata, and not to the data itself, we will not be able to really plumb ambiguity’s ‘twisty darkness’— but I would also agree that ‘preserving the value of the unspoken’ is one of the major challenges for the humanities, especially, as it grapples more deeply with computation and information science.

If you look at where the funding for cyberinfrastructure and the research programs that support it is coming from, you might conclude that not all communities, not all classes, and not all categories are going to be equally well served in its design and deployment. To begin with, and lest we lose sight of this, the largest investment will be from the commercial sector, and it will take the form of developing products, not doing basic research, much less doing education. A knowledge society implies an information economy, and we have already seen that the owners of large caches of information—television and film studios, the recording industries, publishers—are not eager to achieve technical solutions to the problem of restricted access.

In the area of basic research, most of which is now done in universities and colleges, the big dog is health research, which accounts for more than half of federal spending on research: the NIH’s budget in 2004 was around \$28B. By comparison, the NSF’s budget was about \$5.5B (an increase of \$171M over 2003), within which CISE represents about ten percent, or \$584M. By comparison, again, \$584M is about what the largest private foundations give away in a year, in this country—with the exception of the Gates Foundation, which gives away about twice that. Descending the scale of funding and influence, the 2004 budget for IMLS was about \$262M (roughly half of the budget for CISE); the budget for NEH was \$162M (less than the increase for CISE, over 2003); bringing up the rear, the budget for NEA was \$139M—just under half a percent of the budget for NIH. In fact, add the budgets of NEA, NEH, and IMLS together, and you will not quite equal the budget for CISE, which is one of the mid-range budgets in NSF.

‘Cyberinfrastructure’ is more than just hardware and software, more than bigger computer boxes and faster wires connecting them. The term describes new ‘research environments’ in which disciplinary experts, in interdisciplinary teams, supported by specialised computational support staff, have global, instantaneous access to enormous computing resources. And although the redaction of the Atkins report in NSF

presentations subsequent to its publication has tended, in my view, to emphasise only that last point—enormous computing resources—if you read the report itself, you may be struck, as I was, by its emphasis on human resources, on organisations, and on education and training, for example in passages like this one:

This [vision of science and engineering research] involves significant educational dimensions. The research community needs more broadly trained personnel with blended expertise in disciplinary science or engineering, mathematical and computational modelling, numerical methods, visualisation, and the sociotechnical understanding about working in new grid or collaboratory organisations.<sup>10</sup>

Here is another such passage:

Human resources are critical to getting cyberinfrastructure and applications working, keeping them working, and providing user support. In the interest of funding more grants, NSF has arguably under-supported the recurring costs of permanent staff, preferring to focus resources on direct research costs and ‘hard’ or ‘tangible’ assets. In the ACP [Advanced Cyberinfrastructure Program], human resources are the *primary* requirement in both development and operations, and success is clearly dependent on adequate funding both in centers and in the end-user research groups.<sup>11</sup>

What is being said, in these passages, of the importance of discipline-specific computational support in the sciences could also be said of computational humanities or social science. And with respect to the nature of research projects themselves, it would be as true, these days, of the humanities as of science to say that ‘many contemporary projects require effective federation of both distributed resources (data and facilities) and distributed, multidisciplinary expertise, and that cyberinfrastructure is a key to making this possible.’

---

<sup>10</sup> Daniel Atkins et al, *Revolutionizing Science and Engineering through Cyber-infrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure* (January 2003) 17 <[http://www.communitytechnology.org/nsf\\_ci\\_report/](http://www.communitytechnology.org/nsf_ci_report/)>.

<sup>11</sup> Daniel Atkins et al, *Revolutionizing Science and Engineering through Cyber-infrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure* (January 2003) 60 <[http://www.communitytechnology.org/nsf\\_ci\\_report/](http://www.communitytechnology.org/nsf_ci_report/)>.

---

What the Atkins report says—that one could *not* say of the humanities or perhaps even of social science—is that ‘prior investments provide a sound foundation for the ACP.’ In fact, prior investment in cyberinfrastructure for the humanities and social sciences is tiny, by comparison to what it has been in the natural sciences, and that puts us in a rather different position, with rather different needs. In addition to the two ‘resource constraints’ I discussed earlier—intellectual property and privacy rights—we have another resource constraint, and it is a constraint on a human resource, namely those disciplinary computational specialists. Education and training, therefore, must be even higher on the agenda for the humanities and social sciences than they are for computational science or computer science itself. Schools like the one I recently moved to, schools of library and information science, are our best bet for producing those specialists, I believe, and I firmly believe (with Margaret Hedstrom and John King) that libraries are one of the principal places that they will do their work. I also think, though, that we will have to reassert, more generally, the importance of mathematics in general education and in the liberal arts curriculum, beginning as early as middle school and high school. We will need English majors who have a background in logic, who can handle statistics, who do math, if we are going to turn out a generation of disciplinary specialists who can bring the accumulated wisdom of the humanities to bear in computational contexts—perhaps in helping build ontologies for scholarly projects in disciplinary contexts, or building tools for data-mining in the context of humanities research. I have met a few of them at the University of Virginia, and even graduated a couple with PhDs in English who have gone on to tenure-track appointments as humanities computing specialists in English departments; I am meeting and graduating a lot more of them now, at the University of Illinois’ Graduate School of Library and Information Science. These newly minted scholars, some of whom are specialists in disciplines of the humanities, or social sciences, and some of whom are specialists in information science, have arrived at that expertise without abandoning mathematics and logic. Consequently, they have absorbed and naturalised computational methods, and they hunger for more data. Given the necessary resources, they will—I am convinced—find novel ways to bring their disciplines to bear on the uncertainties, the



quandaries, the moral and aesthetic challenges, as well as the practical problems, of ‘the knowledge society.’

Those are some of my starting hypotheses, as chair of the ACLS commission on cyberinfrastructure for the humanities and social sciences. I expect that the work of the commission will test those hypotheses in various ways, and will alter them as a result. I present them here, not as predictions of the commission’s outcomes, but to mark my own starting point, and I would welcome your response to any of the points just raised. During the coming year [2005], the ACLS Commission is charged to:

- Describe and analyse the current state of humanities and social science cyberinfrastructure;
- Articulate the requirements and the potential contributions of the humanities and the social sciences in developing a cyberinfrastructure for information, teaching, and research; and
- Recommend areas of emphasis and coordination for the various agencies and institutions, public and private, which contribute to the development of this cyberinfrastructure.

## CONCLUSION

I will close by observing that there is a kind of ten-year cycle to the sort of thing we are doing. Ten years ago, it was the National Information Infrastructure, and various commissions and committees around that term, that sparked a good deal of the priority-setting and decision-making that set the research agenda for the next decade. The humanities and the arts were a small part of that conversation, and there were some outcomes from that, but I hope this time around the engagement is a more profound one, and I hope the outcomes are more lasting. I also hope that the bridges we build in this process are bidirectional, and encourage collaborations and provocations that finally unite CP Snow’s two cultures, and deconstruct that binary opposition once and for all.<sup>12</sup>

---

<sup>12</sup> For further references, websites and initiatives in this area see: ACO\*HUM (Advanced Computing in the Humanities, sponsored by the European Commission, published in 1999)

## POSTSCRIPT

The report of the American Council of Learned Societies (ACLS) Commission on Cyberinfrastructure for the Humanities and Social Sciences titled ‘*Our Cultural Commonwealth*’ was published in 2006. It proposed that ‘an effective and trustworthy cyberinfrastructure for the humanities and social sciences will have the following characteristics:

- Be accessible as a public good
- Be sustainable

---

<<http://helmer.aksis.uib.no/AcoHum/book/>>; The National Information Infrastructure: Agenda for Action (1993) Ron Brown, Secretary of Commerce and Chair, Information Infrastructure Task Force <<http://www.ibiblio.org/ni/NII-Table-of-Contents.html>>; Humbul Humanities Hub <<http://www.humbul.ac.uk/>>; Voice of the Shuttle <<http://vos.ucsb.edu/>>; H-Net: Humanities and Social Sciences Online <<http://www.h-net.msu.edu/>>; NEH-funded online projects <<http://www.neh.gov/projects/online.html>>; Association for Computers and The Humanities <<http://www.ach.org/>>; Association for Literary and Linguistic Computing <<http://www.allc.org/>>; National Initiative for Networked Cultural Heritage <<http://www.ninch.org/>>; Consortium for Computers in the Humanities/Consortium pur ordinateurs en sciences humaines (coch-cosh) <<http://www.coch-cosh.ca/>>; Association for Computational Linguistics <<http://www.aclweb.org/>>; The American Association for History and Computing <<http://www.theahc.org/>>; Culture, Creativity and Information Technology, Social Sciences Research Council <<http://www.ssrc.org/programs/ccit/>>; History and Geography: Assessing the Role of Geographical Information in Historical Scholarship (2004) <<http://www.newberry.org/hgis/>>; Digital Resources in the Humanities <<http://www.drh.org.uk/>>; Inaugural Conference on Computational Social Science (2003) <<http://socialcomplexity.gmu.edu/5-2003conf/5-2003conf.htm>>; [and see The GIS History Project (1996), <http://www.geog.buffalo.edu/ngia/gishist/>, also Past Time, Past Place: GIS for History (ESRI: 2002)]; ACLS Occasional Paper No 41, *Computing and the Humanities: Summary of a Roundtable Meeting* (1998) <<http://www.acls.org/op41-toc.htm>>; ACLS Occasional Papers No 36, *New Connections for Scholars: The Changing Missions of a Learned Society in an Era of Digital Networks* (1997) <<http://www.acls.org/op36.htm>>; Institutional Models for Humanities Computing <<http://www.kcl.ac.uk/humanities/ccch/allc/imhc/>>; Jahrbuch für Computerphilologie 4 (2002) <<http://www.computerphilologie.uni-muenchen.de/jahrbuch/jb4-content.html>>; *Humanities and Arts on the Information Highways* CNI/ACLS/Getty (1994) <<http://www.cni.org/projects/humartway/>>; Malhotra, Yogesh; Abdullah Al-Shehri and Jeff J Jones (1995) *National Information Infrastructure: Myths, Metaphors And Realities* <<http://www.brint.com/papers/ni/>>

- Provide interoperability
- Facilitate collaboration
- Support experimentation<sup>13</sup>

It further explained that ‘the necessary characteristics outlined above may be thought of as specifications for a humanities and social science cyberinfrastructure. Actually building something that answers to those specifications will require sustained effort and commitment in at least eight areas:

- Invest in cyberinfrastructure for the humanities and social sciences, as a matter of strategic priority
- Develop public and institutional policies that foster openness and access
- Promote cooperation between the public and private sectors
- Cultivate leadership in support of cyberinfrastructure from within the humanities and social sciences
- Encourage digital scholarship
- Establish national centers to support scholarship that contributes to and exploits cyberinfrastructure
- Develop and maintain open standards and robust tools
- Create extensive and reusable digital collections<sup>14</sup>

The report concluded by saying:

We should place the world's cultural heritage—its historical documentation, its literary and artistic achievements, its languages, beliefs, and practices—within the reach of every citizen. The value of building an infrastructure that gives all citizens access to the human record and the opportunity to

---

<sup>13</sup> American Council of Learned Societies, *Our Cultural Commonwealth* (2006)  
<http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf> 27–29

<sup>14</sup> American Council of Learned Societies, *Our Cultural Commonwealth* (2006)  
<http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf> 29–39

participate in its creation and use is enormous, exceeding even the significant investment that will be required to build that infrastructure. The Commission is also keenly aware that in order for the future to have a record of the present, we need legal and viable strategies for digital preservation; considerable investment is now required on that front as well. Investments need to be made on the basis of research, and, in this case, a good deal more research is needed on digital preservation, tools, and uses and users of digital collections, in academic settings and beyond.

But this is only part of the realization that the Commission hopes to leave with readers of this report. In a recent public presentation of the draft findings of this report, the Commission's chair was asked, "If your report were a complete success, what would be the result, five or six years from now?" The answer is twofold. First, if this report's recommendations are implemented, then in five or six years, there will be a significantly expanded audience for humanities and social science research among the general public. A relatively small audience on the open Web will still be a far larger audience than scholars in these disciplines have been able to find up to now in academic bookstores, research libraries, and print journals. Second, if the recommendations of this report are implemented, humanities and social science researchers five or six years from now will be answering questions that today they might not even consider asking.<sup>15</sup> (footnotes omitted)

---

<sup>15</sup> American Council of Learned Societies, *Our Cultural Commonwealth* (2006)  
<http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf> 40