The Preservation of Research Data in a Postmodern Culture

by Chuck Humphrey*

John Curtice, one of the plenary speakers at the 2005 IASSIST conference in Edinburgh, presented time-series evidence showing the emergence of postmodern values. Derived from major attitudinal and value surveys since the 1960's, his research reveals a shift in the locus of personal identity formation from institutions, such as organized religion and political parties, to the "marketplace" of individually established identities. In the postmodern world, everyone supposedly shops for her or his own identity using today's educational systems to compile a uniquely packaged identity through self-actualization. According to this theory, institutions no longer serve as touchstones in determining individual identities.

In his address, Professor Curtice also showed research findings that temper the postmodernist interpretation of today's world. Specifically, many of today's values represent an interaction between individuals and institutions. Nevertheless, institutions are increasingly under attack by postmodern values. These assaults even pose a threat to national data archives and how these archives function in today's societies.

Our national data archives are not immune to fundamental changes within our cultures. After all, data archives preserve one aspect of our cultural heritages, namely, digital evidence of research value. Therefore, one would expect major shifts in culture to result correspondingly in changes in the ways in which the record of our cultures are preserved. What aspects of postmodernism present a threat to our national data archives?

The very nature of the Internet reinforces the image of individualism in today's culture. Everyone can have her or his own domain name and an identity on the Internet for a modest fee. Web technology, including recent Web log (blog) software, has the potential of making everyone a publisher. Napster enabled peer-to-peer music distribution, much to the chagrin of the entertainment industry. In Canada, a temporarily publication ban was imposed by Justice Gomery on testimony before a Commission investigating possible misuses of public funds. Shortly after the ban, this information appeared on the Internet from a site in the United States. As a consequence, the Chief Commissioner partially lifted his earlier ban because the evidence had been widely disseminated on the Internet.[1] These examples illustrate why the Internet is perceived as a great leveler enabling individuals to compete against institutional powers within and outside today's legal boundaries.

From the perspective of data services, the Internet enables us to provide access to larger numbers of users than we have been able to support in the past. Dissemination is more direct and responsive to on-demand access to data resources using the Internet. Our profession sees these as admirable qualities of this technology. Individuals are empowered to retrieve data directly and quickly to their desktops providing researchers a sense of autonomy.

An erroneous corollary of this sense of access-autonomy is the concept that everyone on the Internet is her or his own archivist. We see this arising within the discourse around digital repositories and the idea of "self-archiving." This is an unfortunate choice of words to describe the act of contributing individual works to a database of research publications. A motivating force behind digital repositories has been to increase access to scientific findings more quickly and equitably, sometimes even circumventing traditional channels of peer-reviewed print publications. Some commentators have generalized this self-archiving concept to incorporate research data among the digital objects researchers should contribute to digital repositories.

Self-archiving to me is an oxymoron. The act of archiving involves an institutional commitment to preserve knowledge and culture beyond political and technological changes. In the case of research, data archives represent institutions dedicated to the long-term preservation of data. Ideally, data archiving is a process throughout the life cycle of research and involves the full range of contributors to a research project. While sole investigators still contribute to the overall output of research, increasingly research projects are organized around teams, especially research that is inter-disciplinary, comparative, multi-national and large in scale. Consequently, the idea of an individual being her or his own data archivist runs counter to the way major research is being performed nationally and

internationally today. Furthermore, the process of such research engages many stakeholders, including government granting agencies, universities, researchers, data producers, publishers, libraries and data archives. All of these contributors play a role in the life cycle of research data. One function of a data archive is to identify the custodial relationships among these stakeholders throughout the various stages of the data life cycle. The concept of self-archiving is meaningless in the context of a life-cycle model,

Diminishing the value of institutions in a postmodern world and accentuating the individualistic attributes of the Internet threatens data archives as institutions. Research in Canada conducted in conjunction with the National Data Archive Consultation demonstrated the need for institutional support of data to ensure its long-term access. A study of 100 funded projects by the Social Sciences and Humanities Research Council in Canada between 1977 and 1980 found data for only three of these projects in 2001.2 The data for all three projects had been deposited with the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan.3 This Canadian study demonstrates the level of risk that research data face without an institution responsible for their long-term care.

Just as Professor Curtice found an interaction between institutional and individual values in his research, the challenges facing data archives consist of similar competing values. Consider the example represented by digital repositories. These systems are being built with good intentions to create better access. Unfortunately, the preservation commitment, established practices and protocols of digital repositories are less developed and tend to be based on an end-state model of preservation rather than life cycle. They are being constructed in disciplines without roots in archival science and their proponents have initiated discussions using terms that confound issues of preservation.

IASSIST members need to enter this dialogue to clarify key institutional principles about preservation. We need to educate the developers of digital repositories and the stakeholders in the research community that:

· long-term access is dependent upon preservation, that is, access models that largely ignoring preservation will at best only provide short-term access;

· today's technology cannot replace the professional skills and knowledge of data archivists;

· institutions are necessary to support preservation activities over generations of technology and researchers;

· preserving research data requires the involvement and enduring commitment of all major stakeholders in the research community.

These principles need to be apparent in the data preservation, documentation and citation standards that our profession develops as well as the services that we provide in our local institutions. Furthermore, all disciplines now struggling with new requirements to provide access to research data created through public funds need to by aware of and to embrace these preservation principles. 4 The recent OECD ministerial declaration on this issue is an opportunity for IASSIST members to educate disciplines both within and outside the social sciences about the principles of preserving research data.

What options exist to combat postmodern values threatening our data archives? We must be advocates for data archives individually and collectively through our affiliation with IASSIST and other professional organizations. Internet technology supporting individualism in today's culture should be used to promote institutional solutions for preserving research data. Champions for data archives among the stakeholders in the research community need to be identified and supported. We need to be creative in transforming our national data archives so they remain relevant to the knowledge sector in our societies while ensuring that the functions they fulfill in preserving research data are not sacrificed or diminished. Finally, we need to monitor attitudes toward data preservation constantly and to combat indifference toward institutions with mandates to preserve data.

* Chuck Humphrey is head of data library at University of Alberta. He was president of IASSIST 1991-1995. He can be emailed on humphrey@datalib.library.ualberta.ca

Footnotes
1 Allison Hanes, "Media bans meet the Internet Age," The Gazette: Montreal, April 7, 2005, p. A2.

2 Charles Humphrey, "Preserving Research Data: A time for action" in Preservation of electronic records: new knowledge and decision-making: postprints of a conference, symposium 2003, Ottawa: Canadian Conservation Institute, 2005.

3 The ICPSR, which was created in the 1960's, is an institution dedicated to the preservation of social science research data.

4 OECD Declaration on Access to Research Data from Public Funding, adopted on 30 January 2004 in Paris.