The Long Now: Revolutions in knowledge production and exchange in the 21st Century and the need for large-scale, long-term qualitative research of sociotechnical systems

David Ribes and Geoffrey C. Bowker

Abstract*: Today, knowledge production and sharing are being transformed by sweeping changes in information technology, market relations and patterns of collaboration. The grand challenge for qualitative research of sociotechnical systems is to match such global transformations by moving beyond single-scholar one-off research ventures to communal data sets that can be easily shared between investigators, across generations of scholars and in comparative analyses. This document presents seven steps towards the creation of a facility or institution in support of large-scale, long-term qualitative studies of sociotechnical systems.*

Driven by the dominant forces of our time - globalization, technological revolution and a planet-wide project of sustainability - our systems of knowledge production and exchange are undergoing a profound transformation. Fully imbricated in these crucial areas are changes in the ways in which knowledge communities and work are organized at scale ranging from the local, to the national, to the global. This revolution is changing our world: organizations, policy, and scholarship are only beginning to catch up (Edwards, Jackson et al. 2007). In this proposal we focus on moving forward the agenda of qualitative research of large-scale sociotechnical systems.

The grand challenge to today's qualitative research is to supersede the balkanization of scholars through the creation of a facility supporting long-term and collaborative qualitative data analysis, sharing and curation. Scaling up qualitative sociotechnical research is necessarily an institutional venture. While the particular topics, field sites and methods that make up qualitative research are highly heterogeneous, the underlying challenges we present below are common to all.

Our call for methodological and collaborative innovation is best explained via an analogy in the natural sciences. Today, the ecological sciences are in the midst of a revolution in the scale of their research. Twenty years ago, the average ecologist worked on a hectare patch of land for a short period of time, gathered data over a thirty-year career, published results and then the data were lost. With the creation of the Long Term Ecological Research Network (LTER), the National Science Foundation began to change the nature of research. Today, at a number of sites nationally and in consonance with international projects, ecologists are able to look beyond the scale of a field and timeframe of a career: Ecologists now have the prospect of studying ecology and climate locally, nationally, globally, and over spans of time that match ecological change. Shifting ecological research to a global scale was a matter of creating new organizational forms within the ecosciences (such as LTER). In the last twenty years, new sensor grids have come to cover the oceans, land, sky and space. These technologies did not solve the question of scaling; instead they posed new problems as new streams of heterogeneous data poured into the hands of scientists. Managing and preserving ecological data required a new

organizational form. Standardizing data has proven to be a crucial activity in scaling up the natural sciences, but is not a stand-alone solution. Preserving the *meaning* of data is a human affair, requiring continuous curation.   Whilst preservation has been recognized as an issue in recent NSF reports (Berman, Lavoie et al. 2010), no general response to long-term preservation of these datasets exists in any branch of the sciences; rather there is an conflicted field of partial solutions ranging from supercomputer centers to university libraries.

We advocate a similar revolution in the qualitative study of human social life, transforming our research into a study of what Stewart Brand has called 'the long now' (Ribes and Finholt 2009) . The need for thinking in stretches of years to decades is quite apparent. Paul David's classic study (1989) on the 'productivity paradox' of computing showed that introducing computers into the workplace did not immediately yield the productivity gains promised. In fact, productivity declined for twenty years before moving upwards. The cause, he argued, was that it took about 20 years to 'think' the new technology: to use the dynamo and computer not as bad steam generators and typewriters but to realize some of the potentialities of new ways of working and to allow for the necessary social, cultural, organizational and institutional adjustments to be made.

With the advent of the Internet, we are undergoing a more profound change. We are changing our knowledge generation and expression procedures root and branch. Yet currently we remain bound to the book and article format and to the classic nineteenth century technology of files and folders. With the advent of the printed book, most historians would agree that it took at least 200 years for the generation of the intellectual armature we consider intuitive (the index, the bibliography, the footnote, rules about plagiarism). So qualitative researchers need a form of analysis that is actually responsive to the scale, scope and rhythms of the changes we are studying.

And yet generically, we are caught in the same cycle as the ecologists of yore. When we do get to look at social change, we do so in three to five year projects and at a maximum of three to five sites. There is little accumulation of data across the years (unlike in quantitative social science), particularly across sites of research or across researchers. We reinvent the wheel with each investigation. And yet the hallmark for the advance of social science is accumulation and comparison.

Potential solutions exist. We present seven interlocking steps to meet the challenges for the future of sociotechnical studies, together these make up our vision for an institution supporting long-term and large-scale qualitative research:

- *Scaling up:* Set in place a mechanism for large-scale long-term qualitative research. We have some available exemplars such as the census and the human relations area file.

This requires the creation of research teams which have a persistence beyond the short-term, project by project ways of working we currently employ.

- *Interdisciplinary collaboration:* Sociotechnical phenomena do not rest within the domain of a single discipline or research approach. For example, climate change is a matter of individual action and state policy, of technological innovation and economic reorganization. It is a problem that demands the participation of social science but stretches well beyond it, requiring collaboration with ecological, hydrological, and biological scientists. Modeling climate change still pushes the limits of our computing resources, posing research challenges for computer and information scientists.

- *Comparative approaches:* Comparison across cases is one of the most revealing methods within qualitative research, encouraging the identification of crucial similarities and differences as well as enabling generalization. The key to comparison is sharing data across teams of investigators. This means investing in the creation of comparable data, data that are properly documented to facilitate sharing.

- *Sustainable and shareable data archives:* We must explore ways to federate the data collected over multiple investigative projects. This does not mean that they have to be in lockstep: rather it means that researchers must publish their data alongside articles (increasingly the case in the natural sciences today) in such a way that researchers can discern trends happening beyond their noses, long as these may be. (Significant confidentiality issues exist, and should be addressed through the creation of appropriate consent forms and anonymization procedures).

- *Technological Support:* The infrastructure of qualitative research has not kept up with the ambitions of our research. Tools for collecting and organizing qualitative data remain tedious and intended for small-scale efforts. For example, software tools like NVivo have not yet developed a useful collaborative platform.

- *Integration with quantitative and social-network analytic approaches*: The strengths of qualitative research (detailed, in depth, meaning oriented investigations) must be combined with those of quantitative and semi-quantitative approaches such as network analysis that have the strength of scope and summation. This kind of integration has proven very powerful in the field of history (though the Annalists such as Fernand Braudel); however it is still the exception in social science.

- *Creating new forms of cyberscholarship*: In the social sciences, we continue to use the computer as an unwieldy typewriter. There have been some remarkable experiments, often in conjunction with new media artists (see, for example, the journal *Vectors*): however these remain one-off ventures and generally suffer from a lack of funding to create new forms of expression. When we begin to actively scale up qualitative social science, we will have to deploy the data storage, visualization, and hypertext possibilities of the web. Further, we must as a community develop new tools for text analysis consonant with the availability of electronic data.

We have sketched out above steps toward a research infrastructure. Now we turn our attention to some potential outcomes for the social and behavioral sciences. This is of course a work in process: the whole goal of our process is to generate new questions, new ways of working collaboratively, and new forms of presenting knowledge. This list is indicative of some issues, but does not limit the scope of the infrastructure:

> *Disaster Management:* We are all aware of the toll that natural disasters take on human populations. In the past few years we have seen tsunamis, earthquakes and floods unsettling large populations. However we do not know the 'recovery arc' beyond the dispatch of immediate help and assistance. There are long-term issues with social and psychological responses which arguably occur over generations, but at a rough minimum work on a thirty year time scale. Creating the kind of dataset we envisage will enable us both to detect and develop policy for long-term issues. The comparative data will enable us to uncover commonalities between disasters and to locate solutions which can become invisible to others if tied to just one community (for example, the community of famine researchers).

> *Understanding Nationality and Identity:* Researchers in many fields have looked at the ways in which national and ethnic affiliations interplay with identity over time and according to circumstance (e.g. varying roles of the interplay between church and state). These issues have (with some exceptions) been treated on a one-off basis - the radicalization of some Islamic groups, the development of militant fundamentalism in the United States. However, the processes of affiliation and identity are by their very nature generational socially and life cycle dependent individually. Systematic qualitative study would enable us to gain an understanding of these processes at the appropriate temporal and spatial scales. This could lead directly to new policy frameworks for some of the most thorny issues facing us today.

> *Thinking Information Infrastructure:* We have over the past thirty years been creating a fundamentally new mode for collecting, analyzing and presenting knowledge (on the academic side) and more generally a new mode for the organization of communities and governments. Past studies of changes of this scale (for example, the development of lists; the invention of the printing press) have shown that academic and social processes adjust slowly; the combination of social, political, organizational and institutional change takes roughly two hundred years (and there is no reason to believe that our current changes will be any faster). It would be a boon for all social and behavioral science if we could create long-term comparable datasets which would allow us to trace the ecology of changes we are witnessing. The benefit to society is direct: if we can map out this highly complex ecology, we can work out where 'logjams' (reverse salients) are occurring - for example, changes in attitudes to print in academia require a long term involving tenure processes, true interdisciplinarity and the definitions of an academic career. These are certainly going to happen - gaining an understanding of the processes in action will enable us to develop far better policy frameworks than our current generic firefighting mode.

References:

Berman, F., B. Lavoie, et al. (2010). Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information. San Diego.

David, P. (1989). "Computer and Dynamo: The Modern Productivity Paradox in a Not-Too-Distant Mirror." Center For Economic Policy Research Publication **172, Stanford**.

Edwards, P., S. J. Jackson, et al. (2007). "Understanding Infrastructure: Dynamics, Tension, and Design." Report from "History & Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures".

Ribes, D. and T. A. Finholt (2009). "The long now of infrastructure: Articulating tensions in development." Journal for the Association of Information Systems (JAIS): Special issue on eInfrastructures **10**(5): 375-398.