

Studying the History of Social Science Data Archives as Knowledge Infrastructure

Kalpana Shankar
School of Information and Communication Studies
University College Dublin
Belfield 4
Dublin Ireland
Kalpana.shankar@ucd.ie

Kristin R. Eschenfelder
School of Library and Information Studies
4238 H.C. White Hall
600 N. Park Street
Madison, WI 53706 USA

Greg Downey
School of Journalism and Mass Communication/School of Library and Information Studies
5115 Vilas Hall
821 University Avenue
Madison, WI 53706 USA

Abstract: We map out a new arena of analysis for knowledge and cyberinfrastructure scholars: Social Science Data Archives (SSDA). SSDA have influenced the international development of the social sciences, research methods, and data standards in the latter half of the twentieth century. They provide entry points to understand how fields organise themselves to be ‘data intensive’. Longitudinal studies of SSDA can increase our understanding of the sustainability of knowledge infrastructure more generally. We argue for special attention to the following themes: the co-shaping of data use and users, the materiality of shifting revenue sources, field level relationships as an important component of infrastructure, and the implications of centralisation and federation of institutions and resources. We briefly describe our ongoing study of primarily quantitative social science data archives. We conclude by discussing how cross-institutional and longitudinal analyses can contribute to the scholarship of knowledge infrastructure.

Keywords: social sciences; data archives; institutional sustainability

Acknowledgements: The authors would like to thank the Alfred P. Sloan Foundation, the Irish Research Council, the Wisconsin Alumni Research Foundation, and ASIS&T History and Foundations SIG History Fund for financial support.

Introduction

In this paper, we map out a new arena of analysis for knowledge and cyberinfrastructure scholars: Social Science Data Archives (SSDA). SSDA are global information infrastructures that have been influential in the international development of the social sciences, research methods, and data standards in the latter half of the twentieth century. SSDA provide entry points to understand how fields organise themselves to be ‘data intensive,’ to examine the mutual shaping between specific research disciplines and knowledge infrastructures, and to study the evolution of a field’s valuation of different types of data across time and place. Longitudinal studies of SSDA can increase our understanding of the sustainability of knowledge infrastructure more generally.

We provide a brief history of the link between SSDA and the quantification of the social sciences and then outline the conceptual background of examining SSDA as part of infrastructure studies. We argue for special attention to the following themes: the co-shaping of data use and users, the materiality of shifting revenue sources, field level relationships as an important component of infrastructure, and the implications of centralisation and federation of institutions and resources. We briefly describe our ongoing study of primarily quantitative social science data archives. We conclude by discussing how cross-institutional and longitudinal analyses can contribute to the scholarship of knowledge infrastructure.

A Brief History of Social Science Data Archives and Archiving

We define SSDA at two levels of analysis. First, they are individual data archives with particular institutional missions and physical homes. SSDA are also professional organisations and consortia that promote shared values, standards, and goals via collaborative projects, conferences, and publications. Both levels of SSDA have curated and provided access to quantitative social science data for over fifty years, and both continue to exist in many nations alongside flashier examples of cyberinfrastructure and open data repositories.

SSDA predate both computers and the Internet so their long history provides an opportunity to examine the people, tools, and organisations that constitute infrastructure, and how that infrastructure has adapted over time to change. Further, the existence of SSDA and related professional organizations across nations allow for comparisons of national science strategies in relation to knowledge infrastructure.

Historically, the social sciences have enjoyed quantitative data archives since the 1940s with antecedents dating from the 1920s (Green & Gutmann, 2007). Initially the SSDA movement was driven by household spending surveys, census and political opinion data. For example, in the US, data archives such as the Interuniversity Consortium for Political and Social Research (ICPSR) and the Roper Center for Public Opinion Research began as elections and survey research projects, acquiring machine-readable data in the 1940s. Hoping to usher in a new era of comparative and longitudinal secondary research, their sponsors opened them to the general public in the 1950s and 1960s (Eulau, 1989; House et al., 2004; Scheuch, 2003). In the UK, a committee began work in 1963 to promote sharing of output from government social science surveys, and the UK Data Archive was established in 1966 (United Kingdom Data Archives, 2007). The 1960s and 1970s saw the establishment of over a dozen services and professional associations to coordinate quantitative data collection, to promote data sharing, and to educate students and scholars about quantitative and machine processing analysis methods (White, 1977).

While early data archiving — especially in political science and economics — focused on quantitative data, the rise of post-structuralism and critical approaches in the 1970s led to increased popularity of qualitative data. Non-digital qualitative data (e.g., documents, photos) have long been curated, but the first explicitly digital qualitative data archives often developed in conjunction with existing quantitative data projects. For example, Radcliffe College's Murray Research Archive was founded in 1976 with an explicit multi-method collection development goal but was later merged with Harvard's Institute for

Quantitative Social Science (Altman, 2009). The UKDA's 1994 Qualidata project is often cited as a pioneer in qualitative data archiving and has become part of the larger predominately quantitative UKDS archive (Corti, 2005, 2011; Smioski, 2011).

Despite the prominence and longevity of SSDA, their history has never been told in a critical, synthesising way. SSDA figure surprisingly little in larger histories of the social sciences or knowledge infrastructures. Instead, like many stories of technological progress in the digital age, a thin and instrumental tale of institutional winners has survived in the literature of information sciences and survey methodology. The standard narrative, written almost entirely from the perspective of advocates, leaves out much of the contingency and anxiety involved in trying to build and maintain the technology, funding, expertise, and client base. It tends to represent relationships among archives as entirely positive, overlooking periods of competition and discord. Importantly, it also skirts the thorny question of why scholars might or might not choose to use SSDA in the first place – a question that remains raw as new and often uncurated sources of 'open data' proliferate. A deeper, STS-inflected critique of this standard narrative can contribute to current conversations on the sustainability of knowledge infrastructures because problems experienced by SSDA in the past are parallel to challenges facing newer cyberinfrastructure projects today.

Conceptual Background

Edwards (2010) defined knowledge infrastructures as 'robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds'. Today we easily understand infrastructure as encompassing networked digital data repositories, but pre-computer arrangements that supported the collection and distribution of data via punch cards and tape drives is also infrastructure. But, knowledge infrastructures must also be understood as systems with both social and cognitive elements (Starr, 1987).

The story of the development of SSDA fits into what Porter called 'the history of

quantitative objectivity' and the desire for some social science fields to increase their stature and legitimacy (Porter, 1995: xii; see also Porter, 1986). SSDA activities from the 1940s to the 1970s were motivated in part by the desire of some fields to conduct statistically based comparative and longitudinal analysis to improve the quality of their research claims. Use of these methods arguably raised the prestige of fields in the university environment (Porter, 1995). At a national level, political pressures led funding agencies to emphasise their support for more 'empirical' research focused on 'relevant' or 'practical' outcomes - preferences that favored support for quantitative data (Massey, 2000). Moreover, quantitative methods more easily extend scientific communities of practice across nations (Porter, 1995). The social science data archives movement promised just such a technology of distance to the practitioners of comparative survey and census research in different nations (Rokkan, 1979). The dream of creating a global 'demoscope,' as one sociologist dubbed it in the 1940s, promised a way of 'sampling the facts' across nations and time (Dodd, 1946).

SSDA proponents have long envisioned their institutions to be in the business of promoting sharing and reuse of research data; however, the practicalities of the enclosure and disclosure of data have been contentious in SSDA since their inception. Many scholars have envisioned a technological utopia of data sharing (Willinsky, 2000). However, other scholars have pointed to the complications of creating sustainable 'knowledge commons' and have argued that some enclosure/restrictions on release are necessary for sustainability (Ostrom & Hess, 2006).

A final conceptual framework synthesises the earlier frames as set of all-encompassing 'practices' for organising memory. Bowker (2005) writes that a system of saving information, or 'memory practices' are both 'sequential' in that they declare a starting point for a new set of values and protocols in preserving the past, and 'jussive' in that they necessarily define and even value what can and can't be saved (Bowker, 2005: 228). Further, memory practices are always contested and evolving even as they claim to preserve

unchanging understandings across time and space. Telling the history of SSDA reveals such contingency and jockeying, as new memory practices are developed and deployed in an environment of rapidly changing technologies and constantly shifting relationships.

SSDA and the Evolving ‘Social Sciences’

Across different arenas, calls for rationale and scientific decision making led to increased demand for personnel with the skills to work with quantitative data and its associated methodologies. As Igo (2007) described, 'Professional statisticians, government bureaucrats, academic social scientists, and all manner of planners claimed that survey methods, newly ‘scientific,’ were essential’ for understanding change, for ‘managing a complex industrial society’ (Igo, 2007: 5).

The desire for increased rationality in decision-making, also led to increased support for quantitative data archiving and education. As early as 1929, the Social Science Research Council (SSRC)ⁱ included among its key objectives data preservation, dissemination of ‘materials, methods and results,’ and ‘reproducing basic data’ to increase access and encourage reuse (Carpenter, 2007).

Early SSDA proponents argued for increasing institutional capacity to support scholars and decision makers in using social science methods and quantitative data (Miller, 1989: 152). Miller envisioned a social science data service that would ‘make a massive difference in the development of the research skills of individual scholars’ and provide ‘institutional support... like the equivalent of the traditional scholar’s library and reference service’ (Miller, 1989: 152).

The push for quantified social science research was also related to US federal funding for research and the desire for political approval of research (Solovey, 2013). Solovey (2013: 4) describes the early lack of support for social science research in the National Science Foundation (NSF) during the 1950s as stemming from a ‘scientism’ which saw credibility and

progress as tied to ‘rigorous, systematic, and quantitative investigations’. During Congressional hearings on the founding of the NSF, witnesses from the SSRC ‘all accepted a sharp distinction between objective social research on the one hand and social reform, political ideology, and value-laden inquiries on the other hand’ (Solovey, 2013: 22). Sociologist Harry Alpert, who worked at the NSF from 1953 to 1958, ‘crafted a carefully circumscribed strategy that limited NSF support to the so-called ‘hard-core’ end of the social research continuum’ (Solovey, 2013: 149). It was during this time that Alpert linked such principles of objectivity to the very survey data that the nascent SSDA movement would soon rally around (Alpert, 1956).

Two key questions around labour arose in the evolution of SSDA and their relationship to the social sciences (and these questions persist today). The first concern was related to the education and employment of social science researchers. Would their career prospects be better served if they engaged in the slow and costly gathering of primary data or focused on the more rapid and cost effective use of secondary data? Was choosing to specialize in secondary quantitative data analysis a valid career option for a new social scientist, or did prestige require the collecting of one’s own data? A second question: who should have power over and responsibility in and for data archives? Social science professors collect and use the data from archives and may set collection priorities, but the day to day labor of running archives is not rewarded in most scholarly fields. The field of ‘data librarianship’ embraced the responsibility for data archives and archiving work, but lacked direct connection to the scholarly fields represented in the data and access to resource-level decision makers. Finally, computer expertise was needed to set up systems and define solutions for digital preservation, but ongoing systems maintenance was often not rewarded as innovative work in computing research (Bisco, 1966; Nasatir, 1973).

Co-shaping of use and users

The history of SSDA provides an excellent frame with which to examine the mutual shaping of users and technologies through consumption, modification, disuse or reconfiguration of infrastructure (Oudshoorn & Pinch, 2005). Lack of use (either the use of data or the deposit of data) remains a prominent concern for infrastructure (Meyer and Schroder, 2015) and SSDA have a long history of trying to create and maintain user bases and supporters through adjustment to shifting fashions in social science. Further, in order to encourage development of the social science and their base of users and supporters, most SSDA missions include education to use quantitative methods that rely on archived data. For example, while SSDA were often initially driven by narrower missions to collect and preserve electoral, census and public opinion data, numerous examples exist of how SSDA adjusted collection development goals in response to changing scholarly demand for different types of data (e.g., qualitative, economic, gender/race, health) and then provided outreach and training materials on the newly-acquired data sets. Also, numerous examples exist of SSDA approaching new audiences to develop new users (such as local government officials, government agencies, community colleges, and even secondary schools). Similar strategies have been articulated in other knowledge infrastructures. For example, Ribes and Bowker (2009) note that in the GEON project, the participants' learning and training about the informational dimensions of their science reoriented their stance towards their broader intellectual community.

Material concerns: paying the bills

While information infrastructures are highly diverse, they share many material concerns. One ever-present concern is how to pay the bills. The SSDA community includes organisations with a variety of revenue models: subscription-based, contract revenue, core government funding, and others (Kitchin et al., 2015). It is important that scholars critically examine the business models and revenue streams of infrastructure projects and how they change over time (Eschenfelder, 2014). Studying changes in business models potentially helps explain

concomitant changes in relationships (i.e. user communities, host institutions, funders, competing data services) and technology infrastructure.

A related opportunity for analysis is how to package data to attract funders, users, and depositors. Questions of revenue are directly tied to decisions about how to package data, what types of access or use restrictions to put in place, and what types of services to offer. Because SSDA operate under a variety of revenue models, some openly share data, while others have data sets that are restricted to paying members. Some offer unprocessed data openly, but require membership for access to cleaned data. This diversity suggests that there are a variety of models under which SSDA might design configurations of data and services that attract users, induce data owners to deposit high-quality data, and retain the support of member payments, funders or other financial supporters. Examination of the SSDA field's 40 plus years of design decisions about data products and services provide a highly contextualised analysis point through which to understand the complexities about data access that go beyond simplistic dualisms of completely open and unfortunately enclosed, and to consider the shifting technoscientific, sociotechnical, and institutional forces shaping information infrastructure design decisions more generally.

Lastly, we suggest that understanding business models would allow STS researchers to incorporate theory and empirical findings from cognate areas such as economics (Ostrom, 1990), ecology (Brand & Jax, 2007), organisational theory (Lawrence et al., 2009), and production research (Bhamra et al., 2011) can give potential insights into the creation and maintenance of knowledge infrastructures. The lifespan, or sustainability, of infrastructure is a growing concern as researchers, archive practitioners and funders seek to ensure that resources invested in developing information institutions will have benefits that endure beyond the initial funding period (Crow, 2013; Ember & Hanisch, 2013; National Academy of Sciences, 2014; Maron & Loy, 2011a, 2011b). To contribute to the building of more sustainable infrastructure, scholars have called for greater understanding of the long-term

technological, work organisation, and institutional development challenges faced by developers (Dutton & Meyer, 2010; Ribes & Finholt, 2009). Projects have begun to map out shifting technoscientific, sociotechnical, and institutional demands that influence sustainability of science infrastructure (Ribes & Polk, 2014).

Inter-SSDA relationships and field level infrastructure

Working from core STS prescriptions to examine relationships and networks of interactions, we propose attention to both local and field level infrastructure. While individual archives build and maintain infrastructure to provide access and use of social science data, we argue that networks of SSDA organisations represent an additional and equally important institutional layer of field-level infrastructure that both support and constrain individual SSDA. STS scholars should examine the field-level groups, their relationship to individual organisations, and their role in developing and supporting shared values, common practices and assumptions within a field. For example, professional organisations may recruit and indoctrinate new labor into taken-for-granted values or practices. They also may serve as a platform for shared projects such as standards development, provide a forum to disseminate new ideas, or give individuals a mechanism to develop influence in the community through governance and leadership. Consortia of organisations may pool resources to accomplish larger goals like standards development, make arrangements to divide work, spread field level practices into new nations through financial or other support, and advocate to governments or other funders for support of the broad goals of the field (i.e. the importance of data archiving in general). Importantly, field level organisations also compete with each other for the limited financial and attention resources of their member organisations and individuals working in the field.

Internal histories of the SSDA field are often triumphalist, focusing on the success of collaborative projects such as the Data Documentation Initiative, or the more contemporary

DataPASS project (Gutmann et al., 2009; Vardigan et al., 2008) The field literature provides only hints about the tempestuous nature of inter-archive relationships that have over time included competition for territory (e.g., disciplinary, geographic), disputed use of each others' data, competition for extramural funding and project awards, and negative feelings stemming from perceived dominance of some archives during certain periods. The field literature about international consortia focuses on motivating participation and describing ongoing projects. It does not address or explain periods of lack of activity or unsuccessful initiatives.

Field level organizations are part of infrastructure in that they develop knowledge and practices, perform advocacy, create bridges to related groups, and provide resources and legitimacy. STS scholarship can add insights to infrastructure studies by examining competition between the field level groups for resources and prestige, considering less-successful group projects or periods of inactivity, and critical analysis of who does not participate in these organisations. Examples from the SSSA field include IASSIST, formed in the mid 1970s to represent social science data archive professionals globally (Heim, 1980; Adams, 2006; IASSIST, 2015). IASSIST states a goal of bringing together information professionals with social science researchers and computing specialists (IASSIST), but it is unclear to what degree IASSIST has succeeded in being a bridging infrastructure between these groups.

The field also includes several longstanding consortia of archives. These include the International Federation on Data Archives (IFDO) that coordinates field level goals and practices for SSSA internationally. In Europe, the Council on European Social Science Data Archives (CESSDA) has provided a coordinating platform for European projects.

Historical perspective also aids analysis of networks of organisations as infrastructure. Because participants may be more willing to talk about past conflicts than present tensions, it is easier trace the rise and dissolution of relationships over larger periods of time and it facilitates consideration of how shared values, taken for granted practices, and

assumptions develop and are perpetuated or challenged over time.

Nationalisation/federation of infrastructure:

One analysis point relevant to all information infrastructure studies is whether infrastructure develops in a centralised or federated manner. Centralised infrastructures may provide national-level (or even international) services across multiple organisations in a field. Federated infrastructures include independent instantiations of infrastructure that may coordinate efforts in different ways over time. National research policies and funding patterns can directly influence the degree of centralisation or federation in infrastructure development (Rajabifard et al., 2006), leading to different types of field level coordinating organisations and different types of cooperative and competitive relationships within the field. Is it more effective to centralise infrastructure for a sub-field of research (or even for an entire country), or is it better to have multiple distributed manifestations of infrastructure? Exploring how such decisions are made and under what circumstances can reveal how actors mobilise resources, develop policy, collaborate and compete, and stabilise their institutions.

To give an example, SSDA infrastructure developed differently in Europe and the United States, and in various subfields of the social sciences. In Europe, and in many other parts of the world, national data archives have been common; for example, European nations send only one SSDA representative to organisations like CESSDA and IFDO. In the United States and Canada, there has never been governmental recognition of a national SSDA; currently, seven different US archives are members of IFDO. But at various times stakeholders advocated for a vision of the national approach in North America. Further, at various times different US SSDA jockeyed for implicit recognition in the field as the 'US national data archive.' US social science data archive history however has seen repeated calls for creation of, and dedicated federal support for, a recognised national data archive or archives (Social Science Research Council, 1999; Ember & Hanisch, 2013). But, the fact that

European nations send one representative to CESSDA may hide inter-archive competition for status and resources within nations. For example, while the UKDA has been the official representative to CESSDA, other SSDA have long existed in the UK.

Comparing Social Science Data Archives

We have argued that comparative STS oriented historical studies of social science data archives can provide insight into contemporary knowledge infrastructure concerns. Our own work in this area focuses on the history of five SSDA from the US and Europe and their relationship to each other from inception through the development of web based data access. For each of our sites, the archives' collection was founded with, and has historically primarily been dominated by, quantitative data. Our analysis will examine how each responded to increased demands for archiving of more diverse data types including qualitative data. From each case, we have collected historical documents such as annual reports, strategic plans, budgets and meeting minutes. We have also interviewed current institutional staff and past stakeholders. Our case sites include three of the most long-lived SSDA: the Interuniversity Consortium for Political and Social Research (ICPSR) at the University of Michigan and the UK Data Archive (UKDA) at the University of Essexⁱⁱ and The Roper Center for Public Opinion which recently parted ways with the University of Connecticut. Our remaining case sites are all in relationship to ICPSR, UKDA, or Roper. A more recent data archive, the Irish Social Science Data Archive (ISSDA) at University College Dublin collaborates with the UKDA, the LIS Cross-National Data Center in Luxembourg was established, in part, as an alternative to ICPSR.

To obtain a field level perspective, we have analysed documentation of the professional associations and consortia of data archives including historical materials of the International Association of Social Science Information Services and Technology (IASSIST), the International Federation on Data Archives (IFDO), and the Council on European Social

Science Data Archives (CESSDA). We have traced networks of collaboration or competition between SSDA through analysis of relationships depicted in professional journals. We have also compared conversations about data archiving practices and data archiving labor as a profession in literature representing the voices of social scientists and information professionals. There is opportunity for other historical studies to examine other long-standing archives including for example qualitative data archives and linguistic data archives.

Conclusion

For decades, the social sciences have organised themselves to support long-standing domain data repositories - institutions with ties to specific scholarly fields that take on the mission of collecting, organising, preserving, and disseminating data for the purpose of furthering scholarship. They have drawn on archived data to produce longitudinal findings that would otherwise be impossible, and in recent years, large data sets for the types of prediction that have been en vogue across the disciplines. SSDA are an understudied 'space of convergence', which Chow-White and García-Sancho (2012) define as 'technologically mediated processes of communication. They are the space of flows of people, disciplinary expertise, finance, cultural values, institutional ethics, technology, information, data, and code' (Chow-White & García-Sancho, 2012: 125).

In this paper, we urge greater attention on the part of STS scholars to SSDA as knowledge infrastructures for several interconnected reasons. First, their reach and influence across geographical boundaries are well-documented forerunners of today's cyberinfrastructures; their strategies for maintaining themselves over time have both pragmatic and scholarly implications. Secondly, they provide historical exemplars of how fields organised themselves to be 'data intensive', a persistent call that of course echoes today. SSDA serve as exemplars of how different disciplines have responded to the pervasive call for 'data intensive research' (Levallois et al., 2013). Lastly, SSDA provide a way to study

across cases, time, and place in knowledge infrastructures; the availability of historical documentation, reports, and other working documents affords us opportunities to delve deeply into the interplay of budgets, staffing, and other daily administrative and institutional, often invisible labour, with the work of knowledge production and dissemination.

Lastly, STS inquiry could expand standard narratives of SSDA histories by paying attention to what Jackson (2014) argues is an important but undertheorised of knowledge infrastructures and design: breakdown and repair. We propose that other scholars bring to the study of knowledge infrastructure, as Jackson (2014) eloquently writes, ‘a deep wonder and appreciation for activities by which stability (such as it is) maintained’ (Jackson, 2014: 222). The concept of breakdown and repair cuts across the themes we have described - not just the breakdown and repair of particular technologies and artifacts, but of institutions, relationships, even, potentially, the repair of memory itself. Paying attention to stories of discontinuity and rupture pave the way for understanding the ‘mangle of practice’ (Pickering, 1995). While such analyses may be methodologically and empirically complex, attention to these dimensions of SSDA (and other knowledge infrastructures) -- how difficulties and discontinuities are worked with, worked through, and accounted for – would provide needed symmetry to our understandings of how knowledge infrastructures are created, managed, and ended.

References

- Adams, MO (2006) The Origins and Early Years of IASSIST. *IASSIST Quarterly* 30(3): 5-15.
- Alpert, H (1956) Public Opinion Research as Science. *Public Opinion Quarterly* 20(3): 493-500.
- Altman, M (2009) Transformative Effects of NDIIPP: The Case of the Henry A. Murray Archive. *Library Trends* 57(3): 338-351.
- Bhamra R, Dani S & Burnard K (2011) Resilience: the Concept, a Literature Review and Future Directions. *International Journal of Production Research* 49(18): 5375-5393.
- Bisco RL (1996) Social Science Data Archives: A Review of Developments. *American Political Science Review* 60(1): 93-109.
- Bowker GC (2005) *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.
- Brand FS & Jax K (2007) Focusing the Meaning(s) of Resilience: Resilience as a Descriptive Concept and a Boundary Object. *Ecology and Society* 12(1).
- Carpenter K (2007) Toward a New Cultural Design: The American Council of Learned Societies, the Social Science Research Council, and Libraries in the 1930s. In: Augst T & Carpenter K (eds) *Institutions of Reading: The Social Life of Libraries in the United States*. Amherst, MA: University of Massachusetts Press, 164-282.
- Chow-White PA. & García-Sancho M (2012) Bidirectional Shaping and Spaces of Convergence Interactions between Biology and Computing from the First DNA Sequencers to Global Genome Databases. *Science, Technology & Human Values* 37(1): 124-164.
- Corti, L (2005) Qualitative Archiving and Data Sharing: Extending the Reach and Impact of Qualitative Data. *IASSIST Quarterly*, 29(3). Available at:
<http://www.iassistdata.org/downloads/iqvol293corti.pdf> (retrieved 24.1.2016)
- Corti, L (2011) The European Landscape of Qualitative Social Research Archives: Methodological and Practical Issues. *Forum Qualitative Social Research: Sozialforschung*, 12(3): Art 11. Available at: <http://nbn-resolving.de/urn:nbn:de:0114-fqs1103117> (accessed

30.1.2016).

Crow R (2013) *Sustainability of Open Access Services Report. Phase 3: The Collective Provision of Open Access Resources*. Available from: <http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%2fFiles%2fFiler%2fdownloads%2fOpen+Access%2fSustainabilityServices%2fSustainability+of+OA+Services+phase+3.pdf> (retrieved 30.1.2016)

Dodd SC (1946) Toward World Surveying. *Public Opinion Quarterly* 10(4): 470-483

Dutton WH & Meyer ET (2010) *Enabling or Mediating the Social Sciences? The Opportunities and Risks of Bottom Up Innovation*. In WH Dutton & P Jeffreys (eds) *World Wide Research: Reshaping the Sciences and Humanities* (pp. 165–184). Cambridge, MA: The MIT Press.

Edwards PN (2010) *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: The MIT Press.

Ember C & Hanisch R (2013). *Sustaining Domain Repositories for Digital Data: A White Paper*. Ann Arbor, Michigan. Available at http://datacommunity.icpsr.umich.edu/sites/default/files/WhitePaper_ICPSR_SDRDD_121113.Pdf (retrieved 30.1.2016)

Eschenfelder, KR (2014) *Toward the Inclusion of Pricing Models in Sociotechnical Analyses: The SAE International Technological Protection Measure*. In DL Kleinman & K Moore (eds) *Routledge Handbook of Science, Technology, and Society*. (pp. 157-172) New York: Routledge.

Eulau H (ed) (1989) *Crossroads of Social Science: The ICPSR 25th Anniversary Volume*. New York: Agathon Press.

Green AG & Gutmann MP (2007) Building Partnerships among Social Science Researchers, Institution-Based Repositories and Domain Specific Data Archives. *OCLC Systems & Services* 23(1): 35–53.

- Gutmann M, Abrahamson M, Adams M, Altman, M, Arms, C, Bollen K, Carlson M, Crabtree J
Donakowski D, King, G, Lyle J, Maynard M, Pienta A, Rockwell R, Timms-Ferrara
L, & Young C (2009) From Preserving the Past to Preserving the Future: The Data-
PASS Project and the Challenges of Preserving Digital Social Science Data. *Library
Trends* 57(3): 315-337
- Heim (McCook) K (1980) *Social Science Data Archives: A User Study*. (Unpublished doctoral
dissertation) University of Wisconsin, Madison.
- House JS (ed) (2004) *A Telescope on Society: Survey Research and Social Science at the University of
Michigan and Beyond*. Ann Arbor, MI: University of Michigan Press.
- IASSIST (2015). *About IASSIST*. Available at: <http://iassistdata.org/about/index.html> (retrieved
30.1.2016)
- Igo SE (2007) *The Averaged American: Surveys, Citizens, and the Making of a Mass Public*.
Cambridge, MA: Harvard University Press.
- Jackson SJ (2014) Rethinking Repair. In T Gillespie, PJ Boczkowski & KA Foot (eds)
Media Technologies: Essays on Communication, Materiality, and Society.
Cambridge, MA: The MIT Press.
- Kitchin R, Collins S, & Frost, D (2015) *Funding Models for Open Access Repositories*. Maynooth:
Maynooth University and Dublin: the Royal Irish Academy and Trinity College
Dublin.
- Lawrence TB, Suddaby R, & Leca B (eds) (2009) *Institutional Work: Actors and Agency in
Institutional Studies of Organisation*. Cambridge, UK: Cambridge University Press.
- Levallois C, Steinmetz S & Wouters P (2013) Sloppy Data Floods or Precise Social Science
Methodologies? Dilemmas in the Transition to Data-Intensive Research in Sociology
and Economics. In P Wouters, A Beaulieu, A Scharnhorst, & S Wyatt (eds) *Virtual
Knowledge: Experimenting in the Humanities and the Social Sciences* (pp. 151-182).
Cambridge, MA: MIT Press.

- Massey D (2000) Editorial: Practising Political Relevance. *Transactions of the Institute of British Geographers*: 131-133.
- Maron NL & Loy M (2011a) *Funding for Sustainability: How Funders' Practices Influence the Future of Digital Resources*. Available from:
<http://www.webarchive.org.uk/wayback/archive/20140614201525/http://www.jisc.ac.uk/publications/programmerelated/2011/fundingforsustainability.aspx> (retrieved 30.1.2016).
- Maron NL & Loy M (2011b) *Revenue, Recession, Reliance: Revisiting the SCA/Ithaka S+R Case Studies in Sustainability*. Available from:
<http://sr.ithaka.org/sites/default/files/reports/4.4.2.pdf> (retrieved 30.1.2016)
- Meyer ET & Schroder R (2015) *Knowledge Machines: Digital transformations of the Science and Humanities*. Cambridge, MA: The MIT Press.
- Miller WE (1989) Research Life as a Collection of Intersecting Probability Distributions. In H Eulau (ed) *Crossroads of Social Science: The ICPSR 25th Anniversary Volume* (pp. 147-160).. New York: Agathon Press.
- Nasatir D (1973) An Overview of the Archive Movement. In *Data Archives for the Social Sciences: Purposes, Operations and Problems*. Paris: UNESCO.
- National Academy of Sciences (2014) *Strategies for Economic Sustainability of Publicly Funded Data Repositories: Asking the Right Questions*. Available from:
http://sites.nationalacademies.org/pga/brdi/pga_087151 (retrieved 30.1.2016).
- Ostrom, E (1990) *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge, UK: Cambridge University Press.
- Ostrom E & Hess C (2006) A framework for analyzing the knowledge commons. In C Hess & E Ostrom (eds) *Understanding Knowledge as Commons: From Theory to Practice* (pp.41-81). Cambridge MA: The MIT Press.
- Oudshoorn N & Pinch T (eds) (2005) *How Users Matter: The Co-Construction of Users and*

Technology. Cambridge, MA: The MIT Press.

Pickering A (1995) *The Mangle of Practice: Time, Science, and Agency*. Chicago: University of Chicago Press.

Porter TM (1986) *The Rise of Statistical Thinking, 1820-1900*. Princeton, NJ: Princeton University Press.

Porter TM (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.

Rajabifard A, Binns A, Masser I, & Williamson I (2006) The Role of Sub-National Government and the Private Sector in Future Spatial Data Infrastructures. *International Journal of Geographical Information Science* 20(7): 727-741.

Ribes D & Polk JB (2014) Flexibility Relative to What? Change to Research Infrastructure. *Journal of the Association for Information Systems* 15(1): 287–305.

Ribes D & Bowker GC (2009) Between Meaning and Machine: Learning to Represent the Knowledge of Communities. *Information and Organisation* 19(4): 199-217.

Ribes D & Finholt TA (2009) The Long Now of Technology Infrastructure: Articulating Tensions in Development. *Journal of the Association for Information Systems* 10(5): 375–398.

Rokkan S (ed) (1979) *A Quarter Century of International Social Science: Papers and Reports on Developments, 1952-1977*. New Delhi: Concept.

Social Science Research Council (1999) *Data Infrastructure for the Social and Behavioral Sciences: A Report to the Russell Sage Foundation on the Conference 'Wither Social Science Data and Data Infrastructure' Held at New York University, Nov 11-12*. New York University.

Scheuch EK (2003) History and Visions in the Development of Data Services for the Social Sciences. *International Social Science Journal* 55(177): 385-399.

Smioski A (2011) Archiving Qualitative Data: Infrastructure, Acquisition, Documentation, Distribution. Experiences from WISDOM, the Austrian Data Archive. *Forum*

Qualitative Sozialforschung / Forum: Qualitative Social Research 12(3): Art. 18,

Available from: <http://nbn-resolving.de/urn:nbn:de:0114-fqs1103181> (retrieved 30.1.2016)

Solovey M (2013) *Shaky Foundations: The Politics-Patronage-Social Science Nexus in Cold War America*. New Brunswick, NJ: Rutgers University Press.

Starr P (1987) The Sociology of Official Statistics. In W Alonso & P Starr (eds) *The Politics of Numbers* (pp. 7-58). New York: Russell Sage Foundation.

UK Data Archive (2007). *Across the Decades: Forty Years of Data Archiving*. Essex UK: UK Data Archive. Available from: <http://data-archive.ac.uk/about/archive/decades> (retrieved 30.1.2016).

Vardigan M, Heus P & Thomas W (2008) Data Documentation Initiative: Toward a Standard for the Social Sciences. *The International Journal of Digital Curation* 3(1):107-113.

White HD (ed) (1977) *Reader in Machine-Readable Social Data*. Englewood, CO: Information Handling Services.

Willinsky J (2000) *If Only We Knew: Increasing the Public Value of Social Science Research*. New York: Routledge.

i

☐ References to the Social Science Research Council throughout this paper are to the US-based nonprofit organisation founded in 1923 to advance social science research and not to the ESRC's predecessor in the UK which went by the same name.

ii

☐ The official name of the center is now UKDS (UK Data Service). The UK Data Archive label is retained by the University of Essex to describe the physical facility on the campus. The UKDS is a distributed archival service that includes data and services from other universities, including the University of Manchester. Since our archival research and interviews to date have focused exclusively on the Essex service and our study formally ends at 2002, we are retaining the UKDA terminology for the purposes of this paper.