RESEARCH PAPER

Organizational Resilience in Data Archives: Three Case Studies in Social Science Data Archives

Kristin R. Eschenfelder[1] and Kalpana Shankar[2]

[1] School of Library and Information Studies, University of Wisconsin-Madison, Madison, US
[2] School of Information and Communication Studies, University College Dublin, Dublin, IE
Corresponding author: Kalpana Shankar (kalpana.shankar@ucd.ie)

As public investment in archiving research data grows, there has been increasing attention to the longevity or sustainability of the data repositories that curate such data. While there have been many conceptual frameworks developed and case reports of individual archives and digital repositories, there have been few empirical studies of how such archives persist over time. In this paper, we draw upon organizational studies theories to approach the issue of sustainability from an organizational perspective, focusing specifically on the organizational histories of three social science data archives (SSDA): ICPSR, UKDA, and LIS. Using a framework of organizational resilience to understand how archives perceive crisis, respond to it, and learn from experience, this article reports on an empirical study of sustainability in these long-lived SSDAs. The study draws from archival documents and interviews to examine how sustainability can and should be conceptualized as on-going processes over time and not as a quality at a single moment. Implications for research and practice in data archive sustainability are discussed.

Keywords: Sustainability; longevity; social science data archives; organizational resilience

Introduction

Making data available over time is of growing importance to stakeholders in the research enterprise. Funders and researchers are keen to insure that resources invested in research will have benefits that endure beyond the period of individual research contracts and grants. However, the development and maintenance of trustworthy repositories that are useful, usable, and available for the medium and long-term, requires paying attention not just to the technical infrastructure and means of dissemination of digital and digitized data, but also to the institutions themselves that manage data.

As a result, the data archives community is turning its attention to the sustainability not just of digital data, but also of the data archives themselves as organizations. In addition to developing and implementing tools, standards, and practices that emphasise digital preservation and curation, the community is increasingly acknowledging and addressing the need to understand how the underlying business models that data archives and repositories employ contribute to their sustainability over time (Crow, 2013; Knowledge Exchange 2014; LeFurgy 2009; National Academy of Sciences, 2014; Maron, Pickle & Martin 2013). That said, it is important to ask: What is being sustained, what parts of the organization matter, and can the core functions of a repository be maintained? If it is important to do so, what organizational, technical, funding, policies, and other resources are needed and how do they evolve?

The sustainability of any particular data repository could be assessed in a binary fashion (either the archive is 'sustainable' or it is not) at a given point in time (Downs & Chen 2010). However, such an approach may not be the most useful way of examining how a particular data archive STAYS sustainable and the strategies it employs to do so. The strengths and weaknesses of the business models that archives employ can support or hinder their responses to environmental and internal stressors but also be a generative force. While it may be a cliché, a crisis can be an opportunity for positive growth and change. In the

data archives realm, such opportunities include bringing in new data and new stakeholders, divesting the organization of some components of its business, and assessing value in new ways, to name a few.

In this paper, we explore this central question: how have long-lived successful data archives managed particular 'crisis points' in their efforts to maintain themselves over time? In particular, our goal is to understand how long-lived data archives have remained sustainable by empirically examining what changes they have made to their business models to respond to external and internal threats. We are particularly focusing on the organization's capacity to change – to plan, self-assess, conduct environmental scanning and manage continual change, and staff training to manage change (Shankar & Eschenfelder 2015). An empirical examination of business models and institutional strategies over time is illustrative of how particular approaches have been deployed, how they have evolved, and what the implications are for the organization's goals. Using historical documents from three Social Science Data Archives (SSDA) including annual reports, reports to funders, newsletters, meeting minutes etc. and interviews with former and current staff, we examine how these three SSDA responded to a particular internal or external force that potentially challenged their viability. We conclude with discussion of how these findings could be applied to other kinds of data archives and the importance of conducting such research in a wide variety of data-related institutional settings.

What is Sustainability?

SSDA and similar repositories are infrastructures for scholarly work. There has been significant recent interest in studying data repositories not just from the perspective of design and uses, but as complex networks that integrate people, data, organizations, and material artifacts (Bowker 2010; Edwards et al. 2009). SSDA have not figured significantly in these discussions; nor have discussions of SSDA discussed some of the more difficult issues of maintaining values, mission and resources. Understanding how pre-Internet/pre-computing information infrastructures continue to operate over time can yield insights into other contemporary data projects, including concerns about longevity and sustainability.

A significant challenge in studying sustainability of an organization empirically is a lack of agreement (or in some important contexts, even discussion) about what the term sustainability means. There is a lack of cross-disciplinary discussion, even though the complexity of the topic would lend itself to such discussion. In our effort to construct an appropriate framework for our own work, the authors reviewed a number of models of organizational sustainability from across many different fields (Eschenfelder & Shankar, 2016) and delved more deeply into library and information studies (LIS) literature as the most germane to sustainability as applied to digital libraries and data archives. We identified management and business models as key, but often under conceptualized, aspects of sustainability in the LIS professional literature (Eschenfelder et al. 2016).

Our work suggests three overarching ways of writing and thinking about sustainability. The first, primarily drawn from the LIS literature, is one that might be termed a *practitioner-oriented approach*: written from the experience of working with particular repositories and for an audience of other practitioners/professionals facing comparable issues. Rather than developing an overarching framework for sustainability or linking to an existing theory of organizational sustainability, these contributions provide rich descriptions of how particular sustainability concerns play out in specific contexts without necessarily explaining how sustainability concerns interrelate, and typically without ranking their relative importance. Our past analysis of practicioner-oriented depictions of sustainability found the following major themes: technology (encompassing the techniques and tools of digital preservation), management, finance, disaster planning, mission and values, user relations, and external relations (Eschenfelder et al. 2016).

Another important approach to understanding institutional sustainability suggests framing the term to allow for assessment and evaluation of a particular organization or archive against benchmarks, what we call a *snapshot approach* (Knowledge Exchange 2014; Hamilton 2004). If a data repository meets a set of criteria at a particular level, it could be deemed sustainable. Or, the benchmarks can also be seen as recommended steps in an organization's evolution toward increased sustainability. Importantly, the scoring occurs at a particular moment in time, and thus represents a snapshot of an organization's capacities on particular criteria frozen in time. Assessment against the criteria can be repeated over time to capture more longitudinal picture. An exemplar is the Sustainability Index (SI) that was developed in part to guide fledgling open-information/data institutions toward long term sustainability (Knowledge Exchange 2014). It employs a grid framework and it depicts five stages of development, with five being the most sustainable and one the least. A data repository, it is assumed, will proceed through levels of development from one to five along ten

dimensions on the route to becoming more mature and thus sustainable (funding, business planning, business operations, business development, financial management, technical legal and policy skills, governance systems, and organizational structure and interdependencies). Frameworks like the Sustainability Index are helpful for perspective planning by defining what capacities a repository might wish to develop. It also provides one means of operationalizing the term sustainability in terms of the 5 stages and 10 dimensions (Knowledge Exchange, 2014).

Another example of a snapshot approach is the development of typologies of possible business models. For example, The Digital Repository of Ireland released a report in 2015 that enumerated many of the kinds of business models that open data repositories and archives could adopt (Kitchin, Collins, & Frost 2015). These such a typology provide a helpful presentation of possible business models one might chose at a particular point in time, but they do not explore how or why data organizations change their business models over time.

A third approach might be called *temporally-oriented* models of sustainability; these are least deployed in library and data/information repository contexts. These approaches describe how organizations adjust over time through feedback loops based on organizational learning and environmental scanning. One temporal model that focuses on an organization's ability to adjust to opportunities and challenges in their environment is that of organizational resilience (Bhamra, Dani & Burnard 2011). Organizational resilience as a concept examines how organizations maintain functionality over time by detecting threats and adapting to changing conditions. Resilience is an organization's ability 'to return to a stable state after a disruption.' (Bhamra, Dani & Burnard 2011: p 5376), and it is a function in part of the capacity of an organization to make internal adjustments to cope with disruptions. Organizational resilience depicts these processes as adjustments at a critical juncture with feedback loops as the organization 'learns' and revisits earlier decisions (Burnard & Bhamra 2011). Such a cycle includes detection of threats, ability to respond via adjustments, organizational learning about actions taken via on-going monitoring of the environment.

Since there is no one right approach, we developed our operational definition and working approach to sustainability by drawing on all three. We used all of these approaches to guide what we looked for in our cases. We drew upon the practitioner-oriented approach to understand areas of specific concern to repository practice. The 'snapshot approach' provided a set of expert evaluation criteria. The 'organizational resilience' framework encouraged us to consider the bigger picture of our data archives business model histories by organizing their stories into threats, response to threats, and the resulting organizational adjustments learning.

For our analytical work, we operationalize organizational sustainability as (a) the continued operation of an organization that offers data collections and services, (b) where operation relates to technology, preservation, users, institutional relations, business models, and other facets, (c) in the face of on-going internal and external challenges which may or may not be resolved, (d) where stakeholders recognize continuity in the mission, data sources, and value of the data repository, (e) while keeping in mind that each of those elements might evolve over time in response to (c).

Methodology

This paper is part of an empirical study of one particular set of long-lived data repositories: Social Science Data Archives (SSDA). SSDA have operating for over fifty years (more in some cases), curating and managing data generated from surveys, national census data sets, longitudinal research projects, government agencies, and similar (mostly) quantitative data (Bisco 1996) though they have recently been acquiring qualitative data (Corti 2012). While they are not in general 'open data repositories' as that term is usually used, several of them have open components (Armbruster & Romary 2010). SSDA exist in numerous countries and two of those in our study first developed around the same time (early 1960s) and the third in the 1980s. Although their services and data are now digital, their work predates widespread computing facilities; early SSDA efforts were focused on preserving and delivering data sets that were available on punchcard and tape (Bisco 1966; Geda 1979) Parts of their services are still in the same physical location and even have retained staff for decades. Their similarity provides robust between-case comparison (Yin 2003).

We focused our selection of SSDA on those widely acknowledged in the literature as repositories of machine-readable social science research data (Geda 1979; Bisco 1996) and self-identified as having a specific focus on curating such data. Since our overall topic of interest is organizational sustainability, we selected

SSDA that have been in continuous operation for decades. We also focused on SSDA that had relationships with each other and had archived their own administrative/institutional documents. These three are:

- Inter-University Consortium of Political and Social Research (ICPSR) has long been located at the University of Michigan – Ann Arbor in the USA. We are examining data on ICPSR operations from its founding in the early 1960s to the mid 2000s.
- UK Data Archive (UKDA), long located at University of Essex in the UK, is now part of the larger UK Data Service. We are examining UKDA operations from its founding in 1967 to the mid 2000s.
- LIS Cross National Data Center (formerly Luxembourg Income Study) is located in Luxembourg and also at New York University. We have data on LIS operations from its founding in the mid 1980s.

With permission, the authors gathered administrative documents from all three SSDA including annual reports, reports to funders, contracts, newsletters, meeting minutes, and other records and documents depicting the daily workings of the institution. We limited our analysis up until 2002 for several reasons although we were able to obtain some documents from more recent years. For one, we were granted access to the sites of interest and permission to use archival material provided we were conducting a 'historical' study; some of the more recent issues were deemed too sensitive to include. We chose 2002 as an artificial if reasonable 'historical' cut off; preliminary documentary analysis suggested that 2001/2002 heralded a new set of standards and technologies (the Web in particular) that would easily provide another rich set of studies.

In addition to documentary analysis, we conducted semi-structured interviews with former and current staff at each institution. Ethics clearances were obtained via the Human Research Ethics Committee at (University 1; HS-E-12-53) and the Institutional Review Board at (University 2 SE-2012-0401-CP001). For each staff or former staff member we interviewed, we obtained informed consent and permission to record was obtained. These semi-structured interviews took approximately 45 minutes and were conducted in person and audio recorded. Topics included the individual's perceptions of the organization's history, founding, relationships with other SSDA and other archives, memberships in professional institutions, policy development, and collection development. Most germane to the current project, the interviews also included perceived crisis or opportunity points and how they were addressed and resolved (or not).

Transcripts from the interviews and the documents gathered from the archives were coded for pre-selected themes, but also for other emergent themes of interest. All documents were coded by at least one member of the research team. Project members developed analytical summaries or memos around important themes, and then created cross-study comparative memos to understand how threat, response, and organizational learning occurred over the period of time when the 'crisis' occurred.

We first describe the dominant business models and provide a brief history of our case study sites, derived from data analysis and published institutional histories and interviews. We then analyze our data using the framework of organizational resilience by doing the following:

- Identify crisis/opportunity points to which archives were responding.
- Identify adjustments the SSDA made to their business models in response to the crisis at hand. Specifically what aspects of the business model – the extant arrangement of staffing, products, resources, pricing and customer sets - changed? For the purposes of this article, we were less interested in whether the specific crisis was resolved as much as HOW it was resolved.

Case Studies: Histories and Business Models
ICPSR
The Interuniversity Consortium of Political and Social Research (ICPSR) at the University of Michigan-Ann Arbor began in the early 1960s as an archive for political science surveys. Today, ICPSR maintains over 250,000 research data files in the social and behavioral research with specialized collections in 21 area, including education, aging, and other fields (ICPSR 2016).

Over its long history ICPSR has provided a variety of social science data based on opportunities to acquire data sets and in response to changing fashions in social science research. ICPSR has been extensively involved

in quantitative social science/statistics research and education in multiple ways: curating important data sets, developing tools for data analysis, and training researchers in data use at workshops and intensive short courses. The organization has also been involved in multiple initiatives in standards development and data stewardship.

Core funding from the University of Michigan (as well as space, staff contracts, and other contributions), subscriptions, and contracts are the foundation of ICPSR's revenue, but sponsored research/grants and educational activities also constitute an important source of funds. With the subscription model of ICPSR, colleges and universities pay an annual fee to ICPSR that allows access for authorized users. Similar to a journal subscription model, pricing varies based on research activity of the university and the number of students. On the contract side, ICPSR started serving as a data service for United States federal government agencies in the 1970s and contracts grew as a revenue stream in response to plateaus in subscription income and increasing costs. In this part of the business, agencies pay ICPSR to ingest, prepare and distribute their data – often to the entire world, an audience well beyond the classic ICPSR subscription base.

UKDA

The UK Data Archive (UKDA) is now part of a larger UK Data Service, which includes multiple organizational sites. The UKDA was founded at University of Essex in 1967 as the Social Science Research Council Data Bank. While it is now the lead institution of the larger UK Data Service, UKDA is also an independent entity, and for much of its history had its own funding resources. For the purposes of this current paper it is treated in its incarnation as a standalone service. Today UKDA advertises over 6,000 data collections for research and teaching from the social sciences and humanities (UKDA 2017).

Originally UK social science researchers deposited survey data, but beginning in the late 1970s, UK government agencies deposited national reference surveys, census-derived data products, and other data. Scholars and students in the UK get free access to data. UKDA has also been a recognized leader in archiving qualitative social science data.

Since its inception, UKDA has been funded by the Economic and Social Research Council (ESRC) in the UK, which is one of seven funding councils in the UK. These Councils receive block grants from government agencies including the Department of Education and Science (now the Department of Education). The UKDA developed a five-year funding cycle from ESRC, but in 1977–1986 the funding periods were only two years, introducing more churn costs and uncertainties and difficulty in recruiting and retaining staff. Over time, UKDA has also sought funding from other UK government agencies and granting agencies.

LIS Cross-National Data Center (LIS)

The LIS Cross-National Data Center (Formerly known as the Luxembourg Income Study) is housed in both Luxembourg and New York City. Started in 1979, LIS harmonizes and provides access to cross-national microdata on income, expenditures, and other individual and family financial data. LIS data services currently include multiple years' worth of income microdata from around 50 different countries globally, and multiple years of wealth microdata from fifteen countries (LIS 2017).

LIS Data Center creates agreements with national statistical agencies to receive microdata and then LIS harmonizes to a standard set of variables so that scholars can do valid cross-national comparisons and temporal comparisons. Obtaining microdata from statistical agencies is not always easy, data providers have different comfort levels with providing this data which often has personal identifiers within it, and LIS must negotiate data protection agreements with each participating data provider.

LIS provides access to all scholars in a nation as long as some organization, agency, or ministry in that country has provided funds for a national subscription – typically in three-year tranches. In some nations this is paid by a government education agency, in other cases a national funding agency. For example, the United States' subscription is paid by the National Science Foundation and the United Kingdom's by the aforementioned ESRC.

Points of crisis or opportunity

While the 40 to 50 year history of each case encompasses multiple crises and opportunities, we focus in the next sections on one crisis/opportunity and the reaction of the SSDA to the crisis to maintain 'equilibrium' over time. While the analyses are specific to the cases, they are also exemplars of the types of challenges that all large data repositories face over time in some fashion.

We take this approach for several specific reasons that are illustrative of larger themes in data sustainability. One, we hearken back to our claim that understanding organizational resilience requires analysis over time of crisis, responses, and organizational learning. Our crisis stories illustrate how different aspects of a business model become 'weak' at different points in time and how the SSDA responded. Secondly, they demonstrate a variety of challenges to organizational sustainability and a variety of responses. Here we make another specific claim: that discussing the sustainability of an organization or its business model as a whole buries some specific (often non revenue) challenges. Thirdly, our cases illustrate that some challenges are common across data organizations while others are more specific to particular organizational choices and models. We suggest that it is most illuminating to discuss the sustainability of specific parts of the business model (e.g., current pricing model, relationship to other organizations or institutions) in response to threat.

ICPSR: Threat by technology, response by policy (and technology)

For much of its early history, access to ICPSR data was organized on subscriber campuses through an 'organizational representative' (OR), a local staff or faculty member who had received ICPSR training and served as a de facto champion for ICPSR on the subscriber campus. Before computer networking, ICPSR distributed data via tapes or disks to ORs; individuals on the OR's campus requested the data from the OR via tape or loan of the original disks. Annual subscription fees were tiered and tied to the size of the university or college. Very small schools who could not or would not join as individual members received highly discounted access through a 'federated' pricing model. Each federation had a 'hub' with one OR for the whole federation. The OR received data and documentation from ICPSR and provided access to that data and documentation to the other federation members.

Through the 1970s, ICPSR encouraged the clustering of smaller schools into federations with the understanding that the federation would shoulder the costs of self-service within the federation. This was important to ICPSR because big school memberships were already tapped out and small schools were therefore important to increasing membership revenue. Federations solved two problems: capturing the revenue represented by smaller schools while fulfilling the institutional mandate to spread data use.

In the 1990s, a new suite of technologies threatened the sustainability of the federated model: the development of computer networking and the rapid adoption of File Transfer Protocol (FTP) via Bitnet and similar networks. On the one hand, networking college campuses would seem to provide easier data transfer and access than shipping tapes and disks. However, the same networks engendered three threats to federation: the possibility of unauthorized resharing of ICPSR data, the development and uptake of alternative and free data sources outside of ICPSR's purview, and difficulty in tracking data use.

Practically, computer networking meant that scholars who were not paying members of an ICPSR federation could potentially access ICPSR data via network connection to the hub university that was a paying member. An individual researcher could now more easily get desired data from a colleague at the hub institution who had had the tape drive job run, or who had gotten a disk. With the aid of new technologies, one could FTP the data from that hub university researcher. Of course unauthorized sharing was possible with disks and tapes, but more difficult.

With respect to alternative data sources, there was a concern that individual researchers would bypass ICPSR and make their data available to others directly. For example, internal meeting minutes from a 1994 discussion referred to a Harvard economist who used to give ICPSR his data but was now making his data available to other researchers via FTP himself. While 'open data' seems like a fairly recent concern (indeed, several staff members indicated open data as a potential threat in interviews without fully specifying what open data was), it was clear that ICPSR was concerned quite early on that open data sources could potentially reduce the value of an ICPSR institutional membership.

Use tracking also became problematic. The ORs, by serving as ICPSR reps at each campus, tracked use of ICPSR materials and used that information to justify the expense of membership to their home institutions. With FTP access by end users to ICPSR it was initially harder to track usage and therefore to justify the expense of membership, at least with clear numbers.

In the face of these threats, ICPSR began questioning the value of its federated model and determined the federation price was too low. Management feared that free-rider access via computer networks and FTP would discourage new memberships or potentially lead some schools to drop membership altogether. However, ICPSR acceded to the inevitable and by 1996 FTP was their primary distribution method (The CD-ROM was second).

Throughout the 1990s, ICPSPR considered numerous responses to the networking challenge:

· Considered blocking FTP access to all ICPSR materials for all but hub schools. This would force members of federations to continue to go through their hub. This would be a way to break up federations and force schools to become full members.
· Considered raising prices on federation members, but concerns were raised about losing small school members.

ICPSR ultimately responded by in two ways: policy and technology. First, they rethought the federation model. In 1999, ICPSR declared that federations were only for schools who could not afford to otherwise join. This action was also a direct response to a request from university to join a federation where ICPSR perceived the requesting school should be able to pay for their own membership.

They also developed two technologies to meet the challenge. The first dissociated the networking process from hub schools through the creation of ICPSR Direct, its own networking solution. ICPSR Direct allowed members of all types to FTP directly to ICPSR without going through an intermediary. However, this created new challenges. In 2001, there was discussion of not allowing federation members to use ICPSR Direct because of increased staff burden (a burden formerly assumed by the hub school). ICPSR also developed new mechanisms to track usage from specific campuses using direct FTP in order to still provide usage stats and therefore justify membership expense. This was accomplished by tracking a combination of IP address and individual user registration from home campuses.

In summary, ICPSR faced a technological challenge in the form of computer networking that destabilized a significant component of their business model. They responded with new technologies (their own networking and tracking systems), but also through policy and practice. They negotiated price changes among members of a consortium, developed new policies limiting who could be in a consortium, decided what responsibilities federation hubs should shoulder and what type of access federation members receive.

UKDA: Threats by funding churn, response by augmenting revenue, cutting costs and lobbying funders

A type of crisis common among data repositories is threats to funding. The UKDA 'crisis story' illustrates one particularly widespread problem: short term funding cycles. In the early 1980s, while UKDA had grown significantly in budget, staff size and services, the ESRC remained unwilling to provide longer than a two year funding grant. In 1984, ESRC only provided a one year grant. The UKDA's internal documents state that the archive would like 'indefinite' funding (which it has never received to this day). At the same time, the UK Department of Education and Science argued that it was not appropriate for ESRC to provide permanent funding to a data archive. The two-year grant cycle created significant churn costs as the organization expended resources preparing grants and grant reports and justifying prior performance for new grants. It also became difficult to attract quality staff who were uncertain about the longevity of contracted positions at the UKDA. To add to the funding concerns, in 1984 the host institution, University of Essex, began to complain about the cost of supporting UKDA. At that time, the University of Essex was covering approximately 45% of the archive's costs, but the archive provided numerous data services to the campus.

In response to these threats, over the 1980s and 1990s the UKDA took several actions to develop longer-term sources of revenue and cope with the uncertainties created by ESRC funding cycles:

· Management tried to convince the ESRC to provide more years of funding, justifying UKDA as a 'national resource' that deserved ongoing funding, not just grants. The UKDA wrote a report to its funders comparing its funding situations to other major SSDA in Europe and the US. They argued that government funding of archives was a norm in Europe. As a warning, they described the many complexities of the membership fee model chosen by the ICPSR and Roper Center in the US. Stable core funding in accordance with the European models, they maintained, would allow the UKDA time to diversify its revenue sources. Such diversity could include marketing 'packaged datasets' for use in teaching, the establishment of a consultancy service for data analysis, and investigating the market for a data repository for commercial business. They also invoked a cautionary tale: the 1984 Roper Center bailout by the American Social Science Research Council. Roper had tried to become self-funding but failed which resulted in a rescue operation by the American SSRC. This anecdote implied that it was better to provide secure funding than to clean up the mess of a failed data archive later.

- The Archive instituted usage fees for users outside of the UK Higher Education system, such as local government and international users. They also instituted one-off fees for private researchers to get data; developed tiered data deposit fees for small, medium and large data sets; and added modest handling charges for providing data to users.
- In 1984 UKDA staff submitted a number of grant proposals to diversify their revenue and thus reduce reliance on both the University of Essex and SSRC.
- Also in 1984 institutional documents begin to mention a strategy of reorganizing internal work to emphasize the efficiency of the UKDA, but more importantly communicating said efficiencies to stakeholders. One implementation described in documents was the splitting the work of staff in data preservation from those doing data delivery, so that both teams could focus more specifically on their appointed tasks.
- The archive sought new relationships with other government agencies. The UKDA sought support from the Office of Statistics and the Government Statistical Service for funding or support in lobbying for long term funding because the UKDA was now relieving government departments of the burden of storing their data and servicing outside requests. This reframing of the UKDA core mission would include serving as a trusted broker between government departments and external researchers; saving departments money by storing data and servicing post-survey requests (and saving departments staff costs).

In short, UKDA responded to the common threat of funding churn by augmenting revenue streams through the creation of various use and deposit fees and writing grants. They also responded by changing staffing responsibilities to enhance efficiency and to communicate a rhetoric of efficiency to their many stakeholders. Lastly, they did 'reputation work' by lobbying funders for a different arrangement that provided more certainty over time, but also changed their own goals to become broker between government agencies and end users who wanted government data (and thus were a national resource worth supporting). In 1986, a partial solution was reached. The Department of Education and Science agreed to rolling five year funding structure, but onerous accounting methods limited the newfound freedom acquired through support from DES.

LIS: Threat of data acquisitions difficulties, response by formalization, relationship maintenance, and technology

As noted earlier, the LIS Data Center provides access to harmonized cross-national data on income, expenditures, and other individual and family financial data. The 'valued-added work' of LIS is doing that harmonization to make cross-national comparison possible, since different countries and databases use different data structures and fields to collect similar data. The more countries that are represented in LIS, the richer the possible analyses. However, LIS faced the problem that certain countries with excellent data sources were reluctant to provide it, or even more discouraging, did provide data and stopped providing it. The latter affected the quality of data since longitudinal studies would become almost impossible. Reluctance to deposit happened for a myriad of reasons, such as lack of funding or a national user base, or even changes to internal legal requirements and mandates around data privacy. Staff noted in interviews that many of these issues came down to matters of trust. This lack of trust encompassed LIS itself, LIS's users, and the data archive's privacy and security provisions. As some countries argued, why should an external entity housed in New York, and a vaguely described set of international users, be trusted with precious and vulnerable national-level micro data? Moreover, privacy laws in some nations seemingly restricted transfer of microdata.

LIS's responses over time (interviews with staff suggest that this particular crisis is perennial) have included relationship building, reputation work, internal policy adjustment, and new technology:

- The Director of the archive does a great deal of work creating and maintaining relationships with depositors and helping develop new sources of funding for countries they want to add to the archive, some of whom may not be able to afford subscription fees. Personal relationships are relied upon to build trust. However, these are volatile as champions of LIS leave their institutions, governments change, and ministries and agencies are handed over to new political parties.
- New policies and practices have been designed specifically to develop and maintain trust in LIS as an honest broker. LIS introduced an application form to use data which requires a detailed description of the project for which data is intended, how data will be handled, and explicit promises not to use data for unstated purposes. LIS also created a data 'privacy pledge' and

modified this over time to over time to placate data providers and stay current with evolving data protection and privacy regulations in member countries. Furthermore, they expanded their governance regime to include data provider involvement in governance of the archive.
· LIS developed secure data management software, LISSY, in part to reassure data providers about the security of their data.

Conclusion

In this paper, we have used empirical evidence, drawn from interviews and organizational documents, to identify critical junctures that threatened or weakened parts of business models of SSDA over time. We describe how three such data archives moved through crisis points over time by making technological, policy, and practice changes to their business models. Drawing on the organizational resilience model to guide our discussion, we examined at a very fine-grained level how three Social Science Data Archives (ICPSR, UKDA, LIS) have made internal adjustments in reaction to opportunities and challenges in their environment.

In this paper, we employed the resilience framework and demonstrated its applicability to analyzing data archive sustainability. However, there are two limitations of the resiliency framework that merit discussion. The first is that it gives too much credit to organizations for taking positive action to return to a steady state. It may be that a data archive didn't take positive action or took actions that turned out to be futile or were superceded by other actions. Perhaps the organization takes no action, but nevertheless the crisis is resolved – for example from actions taken by others outside of the organization. Finally, sometimes challenges are not resolved. In this paper we showed several instances in which archives were not able to resolve problems in a robust sense, but rather developed capacities to persist in the face of them.

Secondly, the idea of a 'steady state' presented in the resiliency theory should be questioned. Some resilience models examine how organizations make adjustments to overcome challenges (Chowdhury 2013), but assume they then return to some kind of steady state – that is, organizations make changes they need to make in order to keep providing services and achieving their missions. But what if the organization needs to change its core services and its mission in order to overcome challenges? At what point does the scale of organizational change (to remain sustainable) become so profound that the organization itself is no longer the same? (Birnholtz, Cohen & Hoch 2007).

Two key findings from this study that may seem obvious, but are worth making explicit: No institution is indefinitely sustainable. Sustainability is not a state of grace which one might indefinitely enjoy once having met all the criteria. Rather, everybody has some problems all of the time, and new problems constantly emerge that require adjustments to how an organization operates (Ribes & Polk, 2014; Tsoukas & Chia, 2002). Each of our 'long lived' data archives continually adjusted to changing conditions, sometimes gracefully, sometimes not. Although there are some admirable models for understanding what organizations need to pay attention to in order to remain viable, such models only provide a snapshot and don't necessarily create a mindset for thinking about the need to make improvisational changes over the long term.

Second, organizations are not uniformly sustainable (or unsustainable), but have may have some elements that are more sustainable and some that are less so at any given moment in time (Ribes & Polk 2014). The Sustainability Index suggests this by pointing to ten different dimensions on which digital projects may have more or less sustainable practices (funding, business planning, business operations, business development, financial management, technical legal and policy skills, governance systems, and organizational structure and interdependencies) (Knowledge Exchange 2014). Our case studies showed us how some elements of a data archives business model could be working quite well, while other parts were falling apart.

To present an alternative metaphor for thinking about organizational sustainability, scholars of organizational change suggest thinking of the ongoing functioning of an organization as a *temporary accomplishment*, and suggest we pay attention to what actions actors take to maintain that accomplishment. One piece of imagery used to explain this view is that of a gymnast is trying to stay on a balance beam by making continuous small adjustments and improvisations to keep balanced. The achievement of balance is the temporary accomplishment: the adjustments and improvisations should be the focus of our attention (Tsoukas & Chia 2002). From this view, the fact that an organization is sustainable with regard to x, y or z (e.g., pricing or acquisition of new data) is a temporary accomplishment and we need to tell more stories of how actors succeed (and fail) in maintaining those sustainability arrangements.

Another way of thinking about the sustainability of digital collections is consideration of how data cultures and research methods, and the development of scholarly disciplines interact with data organization sustainability. This is beyond the scope of this paper. Although the data and research cultures of the social sciences are diverse, there are some aspects of the social sciences that bear exploring in the context of SSDA resilience. SSDA began with data sets in limited formats but have grown diverse in scope and scale to include everything from numerical data to films, sound and social media data (Borgman 2015). Some social science data become more valuable with time (such as censuses) while others do not. Ethics, 'packaging' (the data are often useless without accompanying study materials and tools), and the difficulties of obtaining particular data sets are potentially likely to impact the sustainability strategies repositories deploy, but making such connections would require much more empirical study.

Across disciplines in general, the longevity of data organizations and archives is of great concern as significant national resources are being mobilized to create data repositories, cyberinfrastructures, and similar repositories for research data. Stakeholders want to insure they will remain sustainable over time (Anderson 2004; Smith 2001) and are looking to emergent frameworks and cases to explore their options once the initial tranche of funding runs out, but as yet few historical or empirical case studies exist (Palaiologk et al. 2012). The OECD has taken up this concern with a recently convened expert group on research infrastructure sustainability (OECD 2017).

The developers of such initiatives should ask themselves and each other: The sustainability of data organizations in relation to what? Which parts of the organization, and what technologies/funding models/staff resources/policies must be improvised over time to safeguard data? No business model will allow an organization to respond nimbly to all threats. Archives will not always confront the 'same old problem', but rather a combination of new and old problems with new and old solutions. To that end, we suggest talking about the sustainability of an organization or its business model as a whole is less useful for developing policy and practice than drilling down on specific dimensions of business models (e.g., current pricing model) and their responsiveness to internal and external pressures. Lastly, more empirical studies of sustainability are needed to develop creative, evidence-based, robust approaches to a growing problem for data sustainability and data repositories.

Additional File
The additional file for this article can be found as follows:

· **Appendix 1.** Interview Themes and Questions. DOI: https://doi.org/10.5334/dsj-2017-012.s1

Acknowledgements

Competing Interests
The authors have no competing interests to declare.

References
Anderson, W L 2004 Some challenges and issues in managing, and preserving access to, long-lived collections of digital scientific and technical data. *Data Science Journal*, 3: 191–201. DOI: https://doi.org/10.2481/dsj.3.191

Armbruster, C and **Romary, L** 2010 Comparing repository types: Challenges and barriers for subject-based repositories, research repositories, national repository systems and institutional repositories in serving scholarly communication. *International Journal of Digital Library Systems*, 1(4): 61–73. DOI: https://doi.org/10.2139/ssrn.1506905

Bhamra, R, Dani, S and **Burnard, K** 2011 Resilience: The concept, a literature review and future directions. *International Journal of Production Research*, 49(18): 5375–5393. DOI: https://doi.org/10.1080/00207543.2011.563826

Birnholtz, J P, Cohen, M D and **Hoch, S V** 2007 Organizational character: On the regeneration of Camp Poplar Grove. *Organization Science*, 18(2): 315–322. DOI: https://doi.org/10.1287/orsc.1070.0248

Bisco, R L 1996 Social science data archives: A review of developments. *American Political Science Review*, 60(1): 93–109. DOI: https://doi.org/10.2307/1953810

Borgman, C L 2015 *Big data, little data, no data: Scholarship in the networked world.* Cambridge, MA: The MIT Press.

Bowker, G C, Baker, K, Millerand, F and **Ribes, D** 2010 Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In: Hunsinger, J et al *International Handbook of Internet Research.* Netherlands: Springer. pp. 97–117.

Burnard, K and **Bhamra, R** 2011 Organisational resilience: Development of a conceptual framework for organisational responses. *International Journal of Production Research*, 49(18): 5581–99. DOI: https://doi.org/10.1080/00207543.2011.563827

Chowdhury, G G 2013 Sustainability of Digital Libraries: A Conceptual Model. In: Aalberg, T, et al. (Eds.) *Research and Advanced Technology for Digital Libraries.* Berlin Heidelberg: Springer, pp. 1–12. DOI: https://doi.org/10.1007/978-3-642-40501-3_1

Corti, L 2012 Recent developments in archiving social research. *International Journal of Social Research*, 15(4): 281–90. DOI: https://doi.org/10.1080/13645579.2012.688310

Crow, R 2013 *Sustainability of Open Access Services Report. Phase 3: The Collective Provision of Open Access Resources.* Retrieved from: http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%252fFiles%252fFiler%252fdownloads%252fOpen+Access%252fSustainabilityServices%252fSustainabililty+of+OA+Services+phase+3.pdf.

Downs, R R and **Chen, R S** 2010 Self-assessment of a long-term archive for interdisciplinary scientific data as a trustworthy digital repository. *Journal of Digital Information*, 11(1): 1–13. Available at: https://journals.tdl.org/jodi/index.php/jodi/article/view/753 (Last accessed 28 October 2016).

Edwards, P N, Bowker, G C, Jackson, S J and **Williams, R** 2009 Introduction: An agenda for infrastructure studies: Introduction. *Journal of the Association for Information Systems*, 10(5): 364–374. Available at: http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1508&context=jais (Last accessed 21 February 2017). DOI: https://doi.org/10.1017/cbo9780511809842.001

Eschenfelder, K R and **Shankar, K** 2016 *Designing Sustainable Data Archives: Comparing Sustainability Frameworks.* Preliminary results paper at the 2016 iConference. Available at: http://hdl.handle.net/2142/89439 (Last accessed 28 October 2016).

Eschenfelder, K R, Shankar, K, Williams, R, Langham, A, Salo, D and **Zhang, M** 2016 What are we talking about when we talk about sustainability of digital archives, repositories and libraries? Poster presentation at: *Association for Information Science and Technology (ASIST)*, 12 September 2016, Copenhagen, Denmark Available at: http://onlinelibrary.wiley.com/doi/10.1002/pra2.2016.14505301148/full (Last accessed 28 October 2016).

Geda, C 1979 Social science data archives. *The American Archivist*, 42(2): 158–166. Available at: http://americanarchivist.org/doi/pdf/10.17723/aarc.42.2.f9428250uk700x41 (Last accessed 13 February 2017). DOI: https://doi.org/10.17723/aarc.42.2.f9428250uk700x41

Hamilton, V 2004 Sustainability for digital libraries. *Library Review*, 53(8): 392–95. DOI: https://doi.org/10.1108/00242530410556210

ICPSR 2016 About ICPSR. Available at: https://www.icpsr.umich.edu/icpsrweb/content/about/ (Last accessed 28 October 2016).

Kitchin, R, Collins, S and **Frost, D** 2015 Funding models for open access digital data repositories. *Online Information Review*, 39: 664–681. Available at: http://dri.ie/sites/default/files/files/funding-models-open-access-repositories.pdf (Last accessed 28 October 2016). DOI: https://doi.org/10.1108/OIR-01-2015-0031

Knowledge Exchange 2014 Sustainability Index. Available at: https://b2share.eudat.eu/record/147/files/Sustainability%20index%20graphic%20final.pdf (Last accessed 28 October 2016).

LeFurgy, W G 2009 NDIIPP partner perspectives on economic sustainability. *Library Trends*, 57(3): 413–426. DOI: https://doi.org/10.1353/lib.0.0044

LIS Cross-National Data Center 2017 Our Data. Available at: http://www.lisdatacenter.org/our-data/ (Last accessed 13 February 2017).

Maron, N L, Pickle, S and **Marcum, D** 2013 *Searching for Sustainability Strategies from Eight Digitized Special Collections.* Ithaca, NY. Available at: http://www.sr.ithaka.org/wp-content/uploads/2015/08/SR_IMLS_20140129.pdf.

National Academy of Sciences 2014 *Workshop: Strategies for Economic Sustainability of Publicly Funded Data Repositories Asking the Right Questions.* Available at: http://sites.nationalacademies.org/pga/brdi/pga_087151 (Last accessed 21 February 2017).

Organization for Economic and Co-operation and Development 2017 *Strengthening the Sustainability and Effectiveness of International Research Infrastructures.* Available at: https://www.innovationpolicy-platform.org/strengthening-sustainability-and-effectiveness-international-research-infrastructures-oecd-project-0 (Last accessed 21 February 2017).

Palaiologk, A, Economides, A, Tjalsma, H and **Sesink, L** 2012 An activity-based costing model for long-term preservation and dissemination of digital research data: the case of DANS. *International Journal on Digital Libraries,* 12(4):195–214. DOI: https://doi.org/10.1007/s00799-012-0092-1

Ribes, D and **Polk, J B** 2014 Flexibility relative to what? Change to research infrastructure. *Journal of the Association for Information Systems,* 15(1): 287–305.

Shankar, K and **Eschenfelder, K R** 2015 Sustaining data archives over time: Lessons from the organizational studies literature. *New Review of Information Networking,* 20(1–2): 248–254. DOI: https://doi.org/10.1080/13614576.2015.1111699

Smith, A 2001 *Building and sustaining digital collections: Models for libraries and museums.* Available at: https://www.clir.org/pubs/reports/pub100/pub100.pdf (Last accessed 28 October 2016).

Tsoukas, H and **Chia, R** 2002 On organizational becoming: Rethinking organizational change. *Organization Science,* 13(5): 567–582. DOI: https://doi.org/10.1287/orsc.13.5.567.7810

UKDA 2017 About Us. Available at: http://www.data-archive.ac.uk/about/publications (Last accessed 14 February 2017).

Yin, R 2003 Designing Case Studies. In: Yin, R (Ed.) *Case Study Research Methods,* 3rd edition. London: Sage Publications, pp. 19–56.